THE JOHN CREAR LIBR.
RECEIVED
DEC 28 1953

Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
HUMAN RESOURCES RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *Human Resources Research Center*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *Human Resources Research Center*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

The Achievement Motive and Recall of Interrupted and Completed Tasks: J. W. ATKINSON 381
Establishment of an Avoidance Gradient under Latent-Learning Conditions: E. R. STRAIN 391
Proactive Inhibition and Associative Facilitation as Affected by Degree of Prior Learning: S. K. ATWATER 400
Effects of Response Alteration and Different Instructions on Proactive and Retroactive Facilitation and Interference: H. H. MCFANN 405
Factors Influencing Verbal Learning from Films under Varying Conditions of Audience Participation: D. N. MICHAEL AND N. MACCOBY 411
Transfer as a Function of Type and Amount of Preliminary Experience with Task Stimuli: A. E. GOSS 419
Distribution of Practice Prior to Solution of a Verbal Reasoning Problem: A. B. SHAKLEE AND B. E. JONES 429
Amount Set and the Length-Difficulty Function for a Self-Paced Perceptual-Motor Skill: C. E. NOBLE 435
Transfer of Training in Motor Learning as a Function of Distribution of Practice: V. E. MONTGOMERY 440
Retention of Transfer in Motor Learning after Twenty-Four Hours and after Fourteen Months: C. P. DUNCAN AND B. J. UNDERWOOD 445
Simultaneous Contrast as a Function of Separation between Test and Inducing Fields: H. LEIBOWITZ, F. A. MOTE, AND W. R. THURLOW 453
The Locus of Short Duration Auditory Fatigue or "Adaptation": J. D. HARRIS AND A. I. RAWNSLEY 457
Erratum: J. A. TAYLOR AND K. W. SPENCE 462

American Psychological Association

Vol. 46 No. 6 December 1953

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors who request them are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas. Priority in publication is given to articles whose authors assume the full cost of publication.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 (4, Suppl.), 389–449. In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Journal of
Experimental Psychology

VOL. 46, NO. 6 DECEMBER, 1953

THE ACHIEVEMENT MOTIVE AND RECALL OF INTERRUPTED AND COMPLETED TASKS[1]

JOHN W. ATKINSON

University of Michigan

One important obstacle to satisfactory integration of the studies of motivation that have utilized the interruption-of-tasks method has been the lack of an adequate measure of individual differences in strength of motivation. The present investigation was an attempt to determine the effect of strength of achievement motivation on recall of interrupted and completed tasks and to determine whether the measure of the achievement motive developed by McClelland and his coworkers (14, 16) fills the need for a measure of strength of motivation.

Several recent reviews (20), particularly Glixman's (7), Alper's (2), and Rosenzweig's (18), treat the literature in a comprehensive manner. The present experiment attempts specifically to resolve the conflicting results of studies in which either presumed individual differences or experimentally induced changes in achievement-related motivation have been related to differences in recall of incompleted and completed tasks. The results of Zeigarnik (22) and Marrow (12, 13), for example, imply that the greater the motivation to achieve, the greater the tendency to recall more incompleted than completed tasks. The results of Rosenzweig (17, 19), Lewis and Franklin (9), and Glixman (8), however, seem to imply just the opposite relationship.[2]

In the present experiment, differences in recall of interrupted and completed tasks are studied in relation to individual differences in strength of the achievement motive (n Achievement) with different experimental instructions comparable to those of the earlier conflicting studies. Different groups of Ss were exposed to three different instructions prior to performance on 20 paper-and-pencil tasks, half of which were interrupted before completion. The different instructions

[1] This paper is based on a portion of a doctoral dissertation presented to the faculty of the University of Michigan (3). The author is appreciative of the helpful interest and criticism of his committee: Drs. D. G. Marquis (chairman), E. L. Walker, D. R. Miller, T. M. Newcomb, and A. F. Zander. The investigation is part of a larger program of research on achievement motivation supported by the Office of Naval Research. A report of this study was given at the APA meeting, 1951.

[2] See also Alper's (1) earlier demonstration that these opposite trends in selective recall can be related to a number of differences in personality.

were designed to vary the probability that *Ss* would perceive completion of tasks as evidence of personal accomplishment (or success) and incompletion as evidence of personal failure. The method used to obtain an n-Achievement score for each *S* has been experimentally validated and elaborated elsewhere (14, 16).

Procedure

Experimental conditions.—In a *Task Orientation* condition, *E* made no deliberate attempt to create any kind of experimental atmosphere. He simply passed out the task folders after being introduced as Mr. —— and read an instruction for performance of the tasks adapted from Marrow (12). Tasks were timed without calling attention to the fact. He interrupted by saying, "All right, we'll go on to the next one now." After the twentieth task he immediately read instructions for the "test of creative imagination" (measure of n Achievement) and followed the standard procedure in administering this measure (14, 16). After the final story, recall was asked for in the following manner: "Now on the back of your story sheet, you are asked to recall as many of the tasks as you can that you did before the story test. Just jot them down in the order they come to mind, not necessarily in the order they were given. Be descriptive enough so that I will know which one you mean. When you can't think of any more, I'll take your paper." The *Ss* took from 2 min. to a 5-min. maximum in recall. Names and sex were not asked for until *Ss* brought their folders to *E*.

A *Relaxed* and an *Achievement Orientation* condition represent alterations of the basic procedure of the Task Orientation condition, on the one hand in the direction of minimizing the importance of the tasks, and on the other of increasing their importance by making them seem to be measures of highly valued attributes.

In the *Relaxed Orientation* condition, *E* lounged on the desk, joked with students, and in general attempted to create a relaxed atmosphere before being introduced as a graduate student who wanted to try out some tasks. This attitude was maintained in making the following remarks before the task instruction: "I have worked out a series of paper-and-pencil tests that I plan to use with college students in some research later. Right now I am simply trying them out to find out which ones are suitable for my purposes. You don't have to sign your names or anything since I'm not interested in your individual scores. However, I will appreciate your serious cooperation so that we can learn something about the suitability of the tests." From this point on the procedure was the same as Task Orientation.

In the *Achievement Orientation* condition, *E* conducted himself in a serious manner in passing out the folders and giving instructions. He was introduced as Mr. —— who had some tests to give the class. The following preliminary statement was made: "During the war years, psychologists were called on to develop many different kinds of tests in order to select people with high executive capacity, intellectual alertness, the capacity for making quick decisions, and leadership. Now I am going to give you a series of tests in order to compare your scores with those of other college students. Without opening the folder, will you write your name and sex on the outside of the folder. If you know your IQ you might write that under your name and your grade point average for the first semester." After completing the same instruction for the tests used in Task Orientation, he added, "Your work will be interpreted as representing the full extent of your ability, so do your best." The procedure already outlined was followed with these changes: (*a*) *E* made it clear to *Ss* that they were being timed; (*b*) while *Ss* worked on the tasks, *E* walked about the room as if noticing how well or poorly they were performing; (*c*) after the seventh task he said, "Change tests quickly, you need the time"; (*d*) after the tenth task he said, "Some of you are taking a lot of time on these."

Measurement of n Achievement.—The four pictures projected on a screen before the group to elicit imaginative stories for the measurement of n Achievement with their usual code designations were: (*B*) two men working in a shop; (*H*) a boy seated at a desk holding his head; (*A*) the heads of two men (TAT 7BM); and (*G*) boy in foreground with surgical mural behind (TAT 8BM).

Stories were scored according to a modification (B) of the original method (A) which correlates .95 with the original method (16). An *S*'s n-Achievement score represents the frequency of imaginative responses (Need, Instrumental Acts, Anticipatory Goal States, etc.) indicating concern over excellence of performance in his stories. Rescore reliability of the n-Achievement scores for the 83 *Ss* was .93. Scores obtained from Picture *H* were eliminated when another study run concurrently (3) indicated certain inadequacies of the scoring procedures when applied to this picture. The scores of *Ss* on the three remaining pictures, *B*, *A*, *G*, having an estimated equivalent-form reliability of .56 (3) were used in the analysis of results.

The distribution of n-Achievement scores within each experimental condition was divided

as near to the median as possible to provide comparable high and low n-Achievement groups within each condition. The mean n-Achievement score of each high group fell between 9 and 10 and of each low group between 2 and 3.

Tasks.—Folders containing one set of 20 paper-and-pencil tasks modeled after Marrow (12) and MacKinnon (11) in a prearranged order and story blanks for the measurement of n Achievement were distributed at the start of regularly scheduled class periods. Two versions of each task had been constructed. One could normally be completed in 75 sec., the time allotted for each task; the other was rarely completed in that time.

Two sequences of the 20 tasks were randomly determined. Location of 10 short and 10 long versions was also randomly determined. Given two forms (XA and ZA) alike in location of short and long versions but different in sequence of tasks, two other forms (XB and ZB) were constructed reversing the location of short and long versions.

Test folders were distributed in such a manner that *Ss* in adjacent seats would be working on different tasks at the same time and every *S* would have the experience of being interrupted when others near him had finished. Some *Ss* in each section received folders containing only short versions of tasks for another purpose not reported here.

The intended four-way classification of *Ss* according to form of the test was precluded by an unequal distribution of the four forms among *Ss* having high and low n-Achievement scores. Therefore *Ss* were classified only in terms of the location of short and long tasks in the series. Forms XA and ZA were combined and will now be referred to as Form A, and their combined counterparts as Form B. The two forms differ in the distribution of short tasks (easy to finish) in the four quarters of the sequence of 20 tasks. On Form A, the distribution of short tasks was 3, 3, 3, and 1; on Form B, 2, 2, 2, and 4. A fortuitous result of the random placement of short and long tasks is a greater number of short tasks in the early (1–5) and late (16–20) serial positions usually favored in recall on Form B, and long tasks in those locations on Form A.

Subjects.—The *Ss* were 83 male students in ten introductory psychology sections at the University of Michigan in the spring of 1949, distributed among experimental conditions as follows: Relaxed Orientation, 27; Task Orientation, 32; Achievement Orientation, 24. To eliminate any effect on recall resulting from great disparities between the number of tasks completed and interrupted, 17 other *Ss* who completed fewer than 8 or more than 12 tasks were excluded as were 9 others who failed to understand directions, confused the order of tasks, or complained of illness.

Treatment of recall data.—The *Ss* were classified according to strength of motivation, experimental condition, and form of the task-test to allow systematic control of the three variables which might influence recall in analysis of variance. Since the number of tasks completed varied between 8 and 12, the percentage of tasks of each type recalled is the appropriate measure of recall. In order to apply analysis of variance, each *S*'s percentage recall was converted to angles by the *arc sin* $\sqrt{percentage}$ transformation (21, p. 447). Subclass *N*'s resulting from the triple classification of *Ss* were not seriously disproportionate. Analysis of variance by the method of proportional subclass numbers was accomplished following an adjustment for expected *N*'s (21, p. 295; 10).

Results

The mean n-Achievement scores of the Relaxed (4.93), Task (5.63), and Achievement Orientation conditions (5.17) were not significantly different ($F = .22$, $df = 2$ and 80). Evidently the experimental procedures did not differ sufficiently to produce over-all differences in intensity of motivation of the magnitude previously reported when stimulating conditions were more extreme (14).

However, the three experimental orientations did produce differences in recall, which are evident when high and low n-Achievement groups within conditions are separated (Table 1). The *Zeigarnik effect* is represented here as the difference between recall of incompletions and completions (IR–CR). Table 1 shows that both of the conflicting trends of earlier studies occur when *Ss* are grouped according to the strength of their motivation to achieve. The *Ss* high in n Achievement show an increasing tendency to recall more incompleted tasks as experimental instructions increase the probability that completion and incompletion will be perceived as success and failure. Just the opposite is true of *Ss* low in n Achievement; the tendency

TABLE 1

MEAN RECALL OF INCOMPLETED TASKS (IR), COMPLETED TASKS (CR), AND DIFFERENCE IN RECALL OF INCOMPLETED AND COMPLETED TASKS (IR–CR)*

Orientation	Form	High n Achievement					Low n Achievement				
		N		IR	CR	IR–CR	*N*		IR	CR	IR–CR
		Observed	Expected**				Observed	Expected**			
Relaxed	A	5	5.8	43.2	41.5	1.7	7	7.9	54.0	43.3	10.7
	B	6	5.6	43.8	43.0	.8	9	7.7	48.1	41.8	6.3
	A+B	11	11.4	43.5	42.3	1.3	16	15.6	51.1	42.5	8.5
Task	A	9	6.8	49.6	40.2	9.4	8	9.4	49.4	43.3	6.1
	B	5	6.7	45.9	46.6	−.7	10	9.1	45.4	45.1	.3
	A+B	14	13.5	47.8	43.4	4.4	18	18.5	47.4	44.2	3.2
Achievement	A	5	5.1	64.7	44.1	17.6	8	7.0	48.4	46.1	2.3
	B	5	5.0	50.4	48.8	1.6	6	6.9	40.8	46.2	−5.4
	A+B	10	10.1	57.7	46.5	11.2	14	13.9	44.6	46.1	−1.5

* Percentage recall of each *S* was converted to angles by the *arc sin* $\sqrt{percentage}$ transformation (22).
** Adjusted subclass *N*'s for analysis of variance by the method of proportional subclass numbers (10, 22).

to recall more incompletions decreases. The difference between high and low n-Achievement groups in showing the Zeigarnik effect is due to diametrically opposite trends in recall of incompletions between Relaxed and Achievement Orientation. While both groups increase in recall of completions, the high n-Achievement group shows an even greater increase in recall of incompletions and the low n-Achievement group an almost equally great decrease.

Tables 2 and 3 provide a basis for evaluating the significance of these differences. Table 2 is a summary of separate analyses of variance for the IR–CR difference, IR, and CR. Table 3 contains tests of the significance of particular differences.

The Zeigarnik effect (IR–CR).—Differences in Zeigarnik effect attributable to the interaction of Motivation with Orientation and to Form were both significant at the 5% level of confidence (Table 2). Since neither first-order interaction involving Form was significant, their sums of squares were combined with that of the triple interaction term to provide a combined estimate of error having 5 *df*.

The interaction of Motivation with Orientation is elaborated in Table 3. The Zeigarnik effect of *S*s low in n Achievement is significantly greater under Relaxed Orientation than Achievement Orientation, while *S*s high in n Achievement show a near significantly greater Zeigarnik effect under Achievement Orientation than Relaxed Orientation. The only significant difference between high and low n-Achievement groups occurs in the Achievement Orientation condition where the high n-Achievement group shows the greater Zeigarnik effect.

As expected, Form A in which interrupted tasks were located in serial positions favoring recall produced the larger over-all Zeigarnik effect; the over-all mean IR (not shown in Table 1) was 51.1 and the CR was 43.1. On Form B, the IR was 45.6 and the CR was 45.1.

Recall of incompleted tasks (IR).—Table 2 shows that Form and the in-

TABLE 2

ANALYSES OF VARIANCE OF DIFFERENCE IN RECALL OF INCOMPLETED AND COMPLETED TASKS (IR–CR), RECALL OF INCOMPLETED TASKS (IR), AND RECALL OF COMPLETED TASKS (CR)

Source	*df*	IR–CR		IR		CR	
		Mean Square	*F*	Mean Square	*F*	Mean Square	*F*
1. Orientation	2	25.72		50.64		95.81	8.89*
2. Motivation	1	64.18		43.31		1.82	
3. Form	1	1161.99	10.91*	638.86	9.36*	77.43	
4. M × 0	2	620.34	5.83*	662.98	9.72*	2.21	
5. M × F	1	71.66		.68		85.98	67.13**
6. F × 0	2	146.05		103.35		28.84	22.52**
7. M × F × 0	2	84.34		66.90		1.28	
Combined Error							
(5, 6, 7)	5	106.49		68.24		—	
(4, 6, 7)	6	—		—		10.78	

* $p < .05$ using the combined error term.
** $p < .05$ using the triple interaction term.

teraction of Motivation with Orientation also contribute significantly to the variance in recall of incompleted tasks. The first-order interactions involving Form were again combined with the triple interaction as an error estimate. Table 3 shows that under Achievement Orientation, the high n-Achievement group recalls significantly more incompleted tasks than the low n-Achievement group. The high n-Achievement group's recall of

TABLE 3

SIGNIFICANCE OF PARTICULAR DIFFERENCES SHOWN IN TABLE 1

Comparison	Mean Difference	$\sigma_{diff.}$	*t*	*p*
IR–CR				
Relaxed to Achievement Orient.				
High n Achievement	9.96	4.46*	2.23	.10
Low n Achievement	−10.02	3.81	2.63	.05
High vs. low n Achievement under Achievement Orient.	12.71	4.27	2.98	.05
IR				
Relaxed to Achievement Orient.				
High n Achievement	14.16	3.57	3.97	.02
Low n Achievement	−6.41	3.05	2.10	.10
Task to Achievement Orient.				
High n Achievement	9.85	3.44	2.87	.05
High vs. low n Achievement				
under Achievement Orient.	13.02	3.41	3.81	.02
under Relaxed Orient.	−7.55	3.22	2.35	.10
CR				
Relaxed to Achievement Orient.				
High n Achievement	4.20	1.42	2.96	.05
Low n Achievement	3.60	1.21	2.98	.05

* The $\sigma_{diff.}$ and *df* are derived from the error term of the appropriate analysis of variance.

incompletions is significantly greater under Achievement Orientation than under Relaxed or Task Orientation. Under Relaxed Orientation, the low n-Achievement group recalls near significantly more incompleted tasks than comparable *Ss* under Achievement Orientation and near significantly more incompletions than the high n-Achievement group under the same relaxed instructions.

Recall of completed tasks (CR).— Differences in recall of completed tasks attributable to the experimental orientations are significant at the 5% level of confidence[3] and there is no interaction of Motivation with Orientation in this case (Table 2). Table 3 shows that both high and low n-Achievement groups increase significantly in recall of completed tasks from Relaxed to Achievement Orientation. And Table 1 shows that there is not even a suggestion of a difference between the two motivation groups under any experimental condition.

The meaning of the significant interaction of Motivation with Form may be quickly summarized. Even though the short, completable tasks were located in relatively unfavorable serial positions for recall on Form A, *Ss* low in n Achievement recalled as many completions (across orientations) on Form A, 44.1, as on Form B, 44.3. The high n-Achievement group, however, showed the expected difference between forms; their CR on Form A was 41.8 and on Form B, 46.1. The Form with Orientation interaction was the result of a greater increase in CR between Relaxed and Achievement Orientation on Form B (5.0) than on Form A (2.8).

[3] Actually, differences due to Orientation are not significant when tested against the Form with Orientation interaction term alone. With only 2 *df* in the error term, differences due to Form with Orientation are sufficiently large to preclude the possibility of revealing over-all differences due to Orientation. However, when all interactions involving Orientation are combined providing an error term having 6 *df*, the differences between orientations are found to be significant at the 5% level of confidence.

Discussion

A study of the relationship of n Achievement to thresholds for recognition of success and failure words by McClelland and Liberman (15) first suggested that while high n-Achievement scores indicated a positive disposition to excel, lower n-Achievement scores may imply something more than indifference or lack of *positive* motivation to achieve. Their data allow the inference that persons high in n Achievement are predominantly success-oriented and have as their goal the feeling of satisfaction accompanying personal accomplishment while persons lower in n Achievement are more concerned with avoiding feelings of failure. This inference provides a basis for interpreting the distinctly opposite trends in recall of *Ss* classified high and low in n Achievement.

The most clear-cut differences in recall between high and low n-Achievement groups occur in the Achievement Orientation condition, where the probability that completion would be perceived as evidence of personal accomplishment (or success) and incompletion as failure was maximized. The greater recall of incompletions and greater Zeigarnik effect by the high n-Achievement group are consistent with the theoretical expectations and empirical findings of both Zeigarnik (22) and Marrow (12, 13): the greater the motivation to complete the tasks, the greater the relative recall of incompleted tasks.

The recall trends of the high n-Achievement group from Relaxed to Task to Achievement Orientation are also consistent with the findings of Zeigarnik and Marrow. When the situation is deliberately designed to *decrease* perception of performance on the tasks as instrumental to the goal of personal accomplishment (Relaxed Orientation), *Ss* high in n Achievement recall only slightly more

incompletions than completions, and relatively few of either compared to their significantly greater recall when success and failure are at stake (Achievement Orientation). This resembles the behavior of Zeigarnik's "disinterested" *S*s who failed to show any preference in recall. We may take this to mean that motivation to achieve was not engaged by the "relaxed" instructions and that other possible motives, e.g., a feeling of obligation to *E*, intrinsic interest in the tasks themselves, etc. (22, p. 303), were insufficient to produce persistent striving towards completion. However, as the perception of completion as success and incompletion as failure is favored by task- and achievement-orienting instructions, the achievement motive is apparently increasingly engaged. Both the recall of incompletions and completions and the Zeigarnik effect increase from Relaxed to Task to Achievement Orientation for the high n-Achievement *S*s—essentially Marrow's finding.

The *S*s low in n Achievement show just the opposite trend in selective recall, the so-called defensive or anxiety-reducing trend. As instructions increase the probability that completion will be perceived as success, and what is apparently more important to them, that incompletion will be perceived as failure, there is an increase in recall of completions, the predominant trend appearing in Rosenzweig's group data (17); and a decrease in recall of incompletions, a trend accentuated in Glixman's group data (8). Both trends were reported by Lewis and Franklin (9). The decreasing tendency of *S*s low in n Achievement to recall incompletions suggests that they are increasingly motivated by an *unmeasured* fear of failure.

Thus when recall of incompleted tasks is viewed as *instrumental behavior*, the traditional interpretation of the Zeigarnik effect, the differences in recall trends reported for high and low n-Achievement groups support the hypothesis that in achievement situations their goals differ. When the goal is to experience feelings of success and personal accomplishment, then persistence of the interrupted activity in recall and subsequent resumption of it are instrumental to attainment of that goal. When, however, the goal is to avoid feelings of failure, non-recall of past failures and presumably non-resumption of previously failed activities are instrumental to the avoidance of renewed feelings of failure.

Interaction of perception and motivation.—This interpretation leans heavily on an assumption that the achievement motive measured in imaginative behavior becomes a determinant of overt striving only to the extent that a particular performance is perceived as instrumental to the goal of personal accomplishment. The achievement motive is viewed as a latent characteristic of personality which is manifested in behavior only when engaged or supported by appropriate environmental cues. This is the assumption commonly made by learning theorists: environmental cues signify the *occasion* for the performance of previously learned instrumental acts (6, p. 32). A hungry man is more likely to reach for, pick up, and chew an object on a table at which he is sitting if the object happens to be a sandwich than if it happens to be an ashtray.

If instructions that the tasks measure an attribute of personal competence engage achievement motivation to a greater degree than instructions designed to deny the importance of the tasks, then there should be evidence that *S*s worked harder in response to the former. An independent estimate of how hard *S*s were working on the tasks is the number of tasks completed.

The expected increase in mean number of completions between the Relaxed and Achievement Orientation conditions occurred for both high and low n-Achievement groups. The smallest mean number of tasks completed (8.91) was that of the high n-Achievement group under Relaxed Orientation, supporting the hypothesis based on the absence of Zeigarnik effect in their recall, that they were relatively disinterested. The mean number of completions increased to 10.00 under Task and further to 10.60 under Achievement Orientation in a manner consistent with

the assumption that achievement motivation was increasingly engaged by these instructions.

The mean number of completions of the low n-Achievement group increased from 9.56 under Relaxed Orientation to 9.61 under Task Orientation to 10.14 under Achievement Orientation. Under Task and Achievement Orientation, when the achievement motive had been engaged, the number of completions was greater for the high n-Achievement group, reflecting their stronger motivation. In general, the hypothesized interaction of perception and motivation in the determination of instrumental striving is supported by analysis of performance data.

But an important question remains: why did the low n-Achievement group complete more tasks, recall more incompletions, and show a greater Zeigarnik effect than the high n-Achievement group under Relaxed Orientation? The facts imply that the low n-Achievement group was more motivated to complete the tasks than the high n-Achievement group in this condition and furthermore, that they were apparently unconcerned about failure. The assumption that the relaxed instructions minimized the probability of completion and incompletion being perceived as success and failure does not rule out the possibility that the relaxed instructions might have engaged some other motive more characteristic of persons in the low than the high n-Achievement group. There is a reason for thinking that this might have been the case. Zeigarnik observed that, in addition to personal ambition, *a feeling of obligation to E* and intrinsic interest in the tasks were often motives to complete the tasks. A recent study by R. Brown (4) has shown that persons in the middle and low thirds of the n-Achievement score distribution obtain higher scores on the F Scale measure of authoritarian personality than persons in the upper third on n Achievement. While the relaxed instruction de-emphasized personal achievement, it did urge cooperation and suggested that E *would be pleased if* S*s complied with his instructions.* There is, in other words, some basis for thinking that the instruction which produced indifference in the high n-Achievement group might have engaged more motivation to comply in the low n-Achievement group.

Reconciliation of conflicting results.—Can we account for the comparability of the trends of *S*s high in n Achievement with the group results reported by Marrow (12, 13) and the diametrically opposite trends of *S*s low in n Achievement with the group results of Rosenzweig (17), Lewis and Franklin (9), and Glixman (8)? Were there no evidence that Marrow's *S*s as a group were more highly motivated to achieve than those of the other three studies, the present findings would be little more than suggestive. But in light of the implications of high and low n-Achievement scores discussed earlier, there is evidence to support the the argument. Marrow's *S*s were all volunteers; *S*s in the other studies were not. Marrow makes a point of mentioning that "no pressure was exerted to compel attendance . . . to insure a cooperative attitude towards the work" (12, p. 16). Glixman's *S*s, on the other hand, were all "draftees"; his *S*s, members of an introductory course, "were told by the instructor that they were expected to devote a two-hour period to departmental research." In addition, the instructor called off the names of *S*s who were to participate each day "to lend prestige to the calling of *S*s" and discourage "bias of sampling which is possible when *S*s volunteer" (8, p. 228). Subjects in Rosenzweig's "informal group" were employed while those in his "formal group" were "enlisted from the freshman student advisees responsible to the director of the clinic and were personally invited by him" (17, p. 65). Lewis and Franklin's (9) *S*s were apparently unaware that they were taking part in an experiment, having been asked individually merely to help out with some work that had to be done.

If persons low in n Achievement are concerned with avoiding failure, it is unlikely that they would voluntarily place themselves in a test situation in which they might fail. On the other hand, an

appeal for volunteers to take part in some kind of testing could be viewed as a challenge by the person highly motivated to achieve. For him, volunteering is a goal-directed instrumental act. So the suggestion for reconciling the gross contradiction between the trend in selective recall reported by Marrow and those of the other studies considered becomes a hypothesis subject to experimental verification: *Ss* who volunteer are characteristically more highly motivated to achieve than a group of randomly selected or drafted *Ss*.

Since both increases and decreases in recall of interrupted tasks occur with changes in experimental instructions when *Ss* are classified according to strength of motivation to achieve, there can be little hope of reconciling many of the other inconsistencies in experiments with no basis for estimating the motivation of particular *Ss*. However, the promise of the measure of n Achievement for extended theoretical integration of studies utilizing the interruption-of-tasks method is indicated when the results of Cartwright's study (5) of the effect of interruption, success, and failure on attractiveness of activities are considered in light of present findings. His experimental condition was comparable to the Achievement Orientation of the present investigation. Spontaneous remarks following interruption and interviews after the experiment revealed that *Ss* who raised their attractiveness ratings of tasks following interruption and failure anticipated success at the time of interruption or viewed failure as a temporary obstacle to be overcome by subsequent success on the same activity. Cartwright suggests that for these *Ss* "it is possible that one could speak of a 'need for success' " (5, p. 12). Those who reported anticipating failure at the time of interruption or viewed interruption "as an escape from certain failure" (5, p. 5) more often decreased attractiveness ratings following interruption or failure.

A simple experiment run by the author confirms the expectation that the differences noted by Cartwright could be attributed to differences in n Achievement. Nineteen *Ss* performed the same tasks used in the present experiment under achievement-orienting instructions and were interrupted on half of them. They were then asked to look through a test booklet similar to the one they had worked on and choose the five tasks they liked the best. Only one of ten *Ss* low in n Achievement chose more incompletions than completions; five of nine *Ss* high in n Achievement chose more incompletions than completions. The predicted difference between the groups is significant at the 5% level of confidence.

Summary

The interruption-of-tasks experiment was performed with male college students under three different experimental instructions presumed to vary the probability that *Ss* would perceive completion as evidence of personal accomplishment (or success) and incompletion as failure. The *Ss* in each experimental condition were classified high or low in motivation to achieve on the basis of a thematic apperception measure of n Achievement.

When instructions clearly signified that completion meant success and incompletion meant failure, *Ss* high in n Achievement recalled more incompleted tasks and showed a greater Zeigarnik effect than *Ss* low in n Achievement. Recall of incompletions and the tendency to show the Zeigarnik effect increased for *Ss* high in n Achievement as instructions increased the probability that completion and incompletion would be perceived as success and failure. Just the opposite trend occurred for *Ss* low in n Achievement.

Results are consistent with a hypothesis advanced by McClelland and Liberman that n Achievement is essentially positive motivation to experience feelings of accomplishment and success and that lower n-Achievement scores imply relatively greater anxiety about failure. In addition, results were interpreted to mean that the n-Achievement score obtained from thematic apperception stories is an estimate of the strength of latent achievement motivation which is manifested in overt striving to the extent that it is engaged by appropriate environmental cues.

A basis is proposed for reconciling the apparently contradictory implications among several earlier studies concerning the relation of Zeigarnik effect to strength of motivation to achieve, and a subsidiary experiment on attractiveness of interrupted tasks is reported to show the value of an independent measure of individual differences

in n Achievement for theoretical integration of experiments using the interruption procedure.

References

1. Alper, T. G. Memory for completed and incompleted tasks as a function of personality: correlation between experimental and personality data. *J. Pers.*, 1948, **17**, 104–137.
2. Alper, T. G. The interrupted task method in studies of selective recall: a re-evaluation of some recent experiments. *Psychol. Rev.*, 1952, **59**, 71–88.
3. Atkinson, J. W. Studies in projective measurement of achievement motivation. Unpublished doctor's dissertation, Univer. of Michigan, 1950.
4. Brown, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469–476.
5. Cartwright, D. The effect of interruption, completion, and failure upon the attractiveness of activities. *J. exp. Psychol.*, 1942, **31**, 1–16.
6. Dollard, J., & Miller, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
7. Glixman, A. F. An analysis of the use of the interruption-technique in experimental studies of "repression." *Psychol. Bull.*, 1948, **45**, 491–506.
8. Glixman, A. F. Recall of completed and incompleted activities under varying degrees of stress. *J. exp. Psychol.*, 1949, **39**, 281–295.
9. Lewis, H. B., & Franklin, M. An experimental study of the role of the ego in work: II. The significance of task-orientation in work. *J. exp. Psychol.*, 1944, **34**, 195–215.
10. Lindquist, E. F. *Statistical analysis in educational research*. Boston: Houghton Mifflin, 1940.
11. MacKinnon, D. W. *Experimental studies in psychodynamics. A laboratory manual.* Cambridge: Harvard Univer. Press, 1948.
12. Marrow, A. J. Goal tensions and recall: I. *J. gen. Psychol.*, 1938, **19**, 3–35.
13. Marrow, A. J. Goal tensions and recall: II. *J. gen. Psychol.*, 1938, **19**, 37–64.
14. McClelland, D. C., Clark, R. A., Roby, T. B., & Atkinson, J. W. The projective expression of needs. IV. The effect of the need for achievement on thematic apperception. *J. exp. Psychol.*, 1949, **39**, 242–255.
15. McClelland, D. C., & Liberman, A. M. The effect of need for achievement on recognition of need-related words. *J. Pers.*, 1949, **18**, 236–251.
16. McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953, in press.
17. Rosenzweig, S. An experimental study of "repression" with special reference to need-persistive and ego-defensive reactions to frustration. *J. exp. Psychol.*, 1943, **32**, 64–74.
18. Rosenzweig, S. The investigation of repression as an instance of experimental idiodynamics. *Psychol. Rev.*, 1952, **59**, 339–345.
19. Rosenzweig, S., & Mason, G. An experimental study of memory in relation to the theory of repression. *Brit. J. Psychol.*, 1934, **24**, 247–265.
20. Sears, R. R. Personality. In C. P. Stone (Ed.), *Annual review of psychology*. Vol. 1. Stanford: Annual Reviews, 1950. Pp. 105–118.
21. Snedecor, G. W. *Statistical methods*. Ames: Iowa State Coll. Press, 1946.
22. Zeigarnik, B. On finished and unfinished tasks. In W. D. Ellis (Ed.), *A source book of Gestalt psychology*. New York: Harcourt Brace, 1938. Pp. 300–314.

(Received March 30, 1953)

Journal of Experimental Psychology
Vol. 46, No. 6, 1953

ESTABLISHMENT OF AN AVOIDANCE GRADIENT UNDER LATENT-LEARNING CONDITIONS[1]

EDWARD R. STRAIN

Columbus Receiving Hospital, The Ohio State University

The context of the present experiment lies in a more comprehensive experimental analysis of the different kinds of mechanisms mediating the behavioral phenomena rather globally grouped together as transfer of learning. Advances in our understanding of the problems within this area depend upon a more differentiated conceptual specification of the properties of the structural aspects underlying learned modifications of behavior (memory traces, habits, cognitive structures), the learning conditions upon which they are functionally dependent, and upon increasingly precise formulations of the principles relating these to the characteristics of nonidentical performance situations. Such principles may vary in the degree of directness of the relationship required between learning and performance situations (from the relatively direct relationship denoted by primary stimulus generalization or transposition to the less direct relationships involved in secondary stimulus generalization or the role of the fractional anticipatory goal reaction). The specific mechanisms proposed vary also with respect to the essentiality of the role of peripheral responses as contrasted with the operation of purely central mechanisms.

It was the purpose of the present experiment to investigate the occurrence of transfer of avoidance learning under conditions which should yield predicted effects according to a mechanism of transfer suggested by current cognitive theories of learning and which preclude the operation of possible alternative mechanisms.

Although the mechanism to be tested is not put forward as a novel one, it is desirable in the interest of clarity to state as explicitly as is possible at present the basic assumptions from which the predicted results have been derived and which underlie the design of the experiment. The following assumptions are here accepted as a basis for exploring their consequences and are not subjected directly to experimental test.

1. The occurrence of behavior in a given situation will result in a complex of memory traces corresponding to the characteristics of the situation perceived during the behavior. (An animal explores freely without primary reinforcement a compartmentalized runway; later it receives several shocks in one of these compartments without having entered it from another compartment or escaping from it into another compartment.)

2. A sequence of perceptually discriminable situations will result in a functionally interrelated system of such memory trace complexes.

3. Later perception of part of a given situation or one of a sequence of situations will tend to activate the total complex of memory traces or sequences of such earlier established. (This assumption differs from the assumption that the activation of such traces is mediated via an interpolated common [anticipatory] response.)

4. Activation of such serial memory traces may include memory trace complexes not originally established in the same temporal sequence of activity, provided that both sequences have an identical or similar memory trace unit in common.

5. The effect upon behavior of the nonperceptual activation of a memory trace unit (expectation) is similar (in ways to be more precisely defined by future experimentation) to the be-

[1] The writer wishes to acknowledge his indebtedness to Professor Karl Zener of Duke University who directed the research herein reported. The dissertation which this paper reports was submitted in partial fulfillment of the requirements for the Ph.D. degree. It was awarded the 1952 Sigma Xi prize by the Duke University Chapter of Sigma Xi.

havior which would occur to the perception of the situation in which the memory trace was originally established. (The introduction of an animal into a compartment, other than the one in which shock was experienced, will result in avoidance of the shock compartment without the animal's perceiving it.)

6. It is further assumed (and this is the only novel assumption made) that the degree of directness of the functional connections between the units of the trace system is a determinant of the probability of activation of the given memory trace, and of the strength of its effect upon present behavior. (The magnitude of the effect will be an inverse function of the number of actual intervening compartments of the runway between the shock compartment and the compartment into which the animal is introduced.)

The present experiment consists, then, in an attempt (*a*) to realize concretely, as indicated schematically in the parenthetic clauses above, the essential conditions specified in the set of assumptions as sufficient to mediate transfer of learning; (*b*) to do this under conditions precluding explanation of the predicted transfer on the basis of alternative transfer mechanisms; and (*c*) to investigate the functional effect of variation in directness of memory trace connections as a condition of magnitude of transfer effect.

Method

Apparatus

Runway.—The runway used in the experiment was an elongated wooden alley divided by swinging doors into six equal-sized, removable compartments. The compartments, constructed of ¼-in. wood, were (internal dimensions) 17 in. long, 5 in. wide, and 5 in. high. They differed in their visual properties by variations in (*a*) the painted patterns on their sides, (*b*) the floor construction, and (*c*) designs on the swinging doors. The runway was constructed so that the four inside compartments could be shifted in position within the sequence, the compartments being stabilized by braces. These were systematically shifted in order to achieve equality in the average degree of stimulus similarity between the end compartment in which shock was administered and each compartment step. By "compartment step" is meant the position of any of the four inside compartments in relation to the shock compartment, and is to be distinguished from the designation of the actual compartments themselves. The actual compartments were labeled "shock compartment" (end compartment), A, B, C, D (middle compartments), and "neutral compartment" (opposite end compartment). Regardless of the changes introduced by shifting the four middle compartments, the compartment constituting one end of the runway will hereafter be designated as "shock compartment," the adjacent compartments in order I, II, III, IV, and "neutral compartment." Thus, the first compartment step means the position next to the "shock compartment," etc.

One side of each swinging door, separating the compartments, was white and the other black, and the doors were ordered in the runway so that a compartment's two doors were either both white or both black. A surgical rubber strip along the door sides insured a tight fit. The top of each compartment was a ground-glass plate, sliding on wooden runners and movable in either direction across the width of the compartment. The glass was translucent; it prevented *S* from perceiving extracompartment contours while (because *S* was only 1 to 2 in. below it) permitting *E* to observe its behavior.

Although the compartments were designed to be readily discriminable, an effort was made to keep any one from being perceptually outstanding. The sides of each were painted in different black and white patterns which, however, provided no directional cues, and the floor of each consisted of brass-covered wooden sections varying in width and pattern. The door designs, which provided *S*s directional cues in getting to other compartments, were of different shapes.

Twelve differing door designs approximately 1¼ in. in size were constructed of adhesive tape to make them transferable. A pair was assigned to each compartment step and remained in that relative position regardless of the actual compartment location. A design also was placed on the interior of each end compartment door. Since the door designs were the only directional cues provided, uniformity of discrimination was insured by using a standard pair (for both training and test trials) in the compartment into which *S* was introduced. These standard designs, a circle and a Greek cross, replaced those otherwise assigned to that compartment step. Possible preferences between these two designs were neutralized by placing the cross on the door leading toward the shock compartment for half of the *S*s in each group, and the circle for the other half.

Shock compartment.—In order to eliminate any olfactory cues from shock administration, the shock compartment was replaced on the shock trial by a duplicate compartment (with an escape-proof door), which in turn was replaced

by the original for the test trial. Alternating current of 110 v. at 60 cycles passed through five 15,000-ohm resistors connected in series to six brass strips (½ in. wide) that ran lengthwise and formed the compartment floor. The arrangement made it impossible for *S* to escape the electric shock by standing or crouching.

Experimental room.—The experiment was conducted in a dual chambered room. The runway was located in the inner, relatively soundproof chamber, and *E* observed *S*s from the outer darkened observation chamber. In addition, further precautions were taken to control extrarunway cues by constructing a smaller area immediately surrounding the runway. Two layers of white sheeting were stretched from wall to wall, forming an artificial ceiling 7 ft. above the floor and about 11 ft. long. At each end sheeting was hung to within 2 ft. of the floor, with a slit in one end allowing for access to the experimental region so formed. Sheeting also was hung along the side walls. Over the partition wall a dull gray screening was placed to obscure the small remaining part of the observation windows.

To achieve a uniform brightness, three 100-w. lights were equally spaced above the artificial ceiling; these were shielded from below to reflect the light from the chamber ceiling. Further diffusion was obtained by the cloth layers forming the artificial ceiling. At each end, two lights were placed on either side so that the lighting intensity of the ends approximated that of the ceiling.

Subjects and Experimental Groups

The *S*s were male albino rats procured from the Albino Farm Colony as required during the experiment. The ages at the beginning of the experimental procedure were approximately 140 days. The *S*s were kept in individual cages in the colony room.

The major variable, as indicated in the problem formulation, was the number of compartments intervening between the shock compartment and the compartment into which *S* was introduced on test. Twenty *S*s served in each of four experimental groups. All *S*s of Group I were introduced into Compartment I, the compartment adjoining the shock compartment. The *S*s of Groups II, III, and IV were introduced into Compartments II, III, and IV respectively.

Rats were run in batches of approximately 20 *S*s, each batch divided randomly among the four experimental groups in the attempt to secure equal representation of each group in each batch. Because some *S*s were eliminated after the fourth training trial (see *Runway training Trials 1–6*) equal representation of all four groups at each test trial was not always possible. Also five *S*s which did not leave the compartments into which they were introduced on test were discarded, since the relevant comparisons involved directions of actual exit. Each group had at least one *S* that behaved in this way. It was thus necessary to run an additional experimental batch to make up for these discarded *S*s.

Procedure

Handling and feeding.—The *S*s were handled for seven days before the beginning of the experiment. This was done for a few minutes just preceding the feeding period which was on a 24-hr. schedule. This 24-hr. feeding rhythm continued throughout the experiment. The food was a moist mixture of Purina Dog Chow mash and water. Approximately 65 gm. of this mixture were available to each *S* every 24 hr. for a 1-hr. period. A yeast tablet was given after feeding, and water was present in the cages at all times.

Runway training Trials 1–6.—Daily for five days following the seventh day of the feeding rhythm each *S*, at 21–24 hr. of food deprivation, was given one training trial consisting of 300 sec. of free exploration of the runway. On the sixth day (on which final training, shocking, and testing occurred), the free exploration period was limited to 150 sec. in order to equalize the duration of final training and test trials.

In order to control as many cues as possible, considerable attention was given to the method of bringing *S*s into the experimental region, introducing them into the runway and removing them from it. The *S*s were carried individually into and out of the region in a cage covered with a large black cloth. When introducing *S* into a compartment, all of the glass covers of the remaining compartments were closed. The *S* was introduced quickly into the open compartment and the glass top closed rapidly. During this operation, as well as during removal, there was a period of about 2 sec. when *S* was exposed to the experimental region. The *E* always stood on the same side of the table for introduction and on the opposite side for removal.

Following the training trials, *S* was detained in a neutral cage in the observation chamber for about 10 min. This was done to minimize any reward value provided by association of removal from the runway with return to the home cage.

In order to insure an adequate background of runway learning for all *S*s at time of test, *S*s that were not moving into all the compartments by the fourth training trial were eliminated. On this basis 64 *S*s were rejected during experimentation.

Shock trial.—After the sixth training trial *S* was kept in a neutral cage in the observation

chamber for 5 min. Then it was put directly into the duplicate shock compartment, and a 3-sec. shock was administered after 30 sec. A second 3-sec. shock followed after another 30 sec. Within 5–10 sec. *S* was removed to a neutral cage and taken to the observation chamber where it remained until tested approximately 15 min. later. On the shock trial the runway had the same orientation with respect to the room as on the sixth training trial (information regarding runway rotation in *Other controls*). About midway through the detention period *S* was picked up and put into a duplicate neutral cage to lessen any tendency to avoid *E*'s hand.

Testing.—Following the 15-min. period after the shock trial, *S* was introduced into the same compartment into which it had been consistently introduced during training. It was left in the runway for 150 sec. and its behavior recorded.

Other controls.—In addition to the precautions taken to minimize the presence of extrarunway cues, the following procedural controls were included in the experimental design. (*a*) To avoid the possibility of consistent association of the shock compartment with one end of the room, the runway direction was reversed for each training trial. (*b*) In each group, one-half of the *S*s were introduced facing one direction and the other half toward the opposite, on the initial training trial. To control the possibility of *S*s' developing a tendency to run toward one end compartment, each *S*'s orientation on introduction was regularly reversed from day to day during the training trials. (*c*) For one-half of each group, the runway position was *reversed* after the shock trial, and the test trial orientation was *opposite* that of the last training trial. These *S*s are referred to as the "different" class, in contrast to the "same" class for which the runway remained in the *same* orientation and whose orientation on introduction was the *same* for the test trial as for the sixth training trial.

Behavior record.—During the training trials and the test trial the time spent in each compartment was entered on a data sheet which was constructed to allow *E* to record the time *S*s spent in each compartment in a pattern which represented their movements while in the runway. Recorded behavior consisted of the primary index of avoidance, i.e., the initial direction of exit from the compartment into which *S* was introduced on test. In addition, latency of exit from this compartment, number of compartments initially traversed before reversal of direction, the cumulative time in each compartment during the 150 sec., and frequency of intercompartment movements were recorded or computed. On shock and test trials, an assistant recorded such additional behavior as urination and defecation, conflict indices such as approach-withdrawals at compartment doors, and dashing or hunching activities.

Results

The most direct deduction from the initial assumptions with regard to behavior occurring in the single compartment in which *S*s were introduced on test concerns changes in the proportion of direction of exit toward or away from the shock compartment. According to Assumption 5, this change should correspond to the behavior which would occur at the actual perception of the shock compartment.

Evidence of transfer of avoidance behavior within the compartment into which introduced on test.—The data concerning initial direction of exit from the test compartment is given for the four experimental groups in Table 1. As measured by a *t* test, the proportion of *S*s initially leaving the compartment in the direction away from the shock compartment (53 out of 80) is statistically significant at the 1% level of confidence. This is consistent with the derivation from the assumptions.

Evaluation of the possible effects of extrarunway cues.—Before it can be concluded that these transfer effects resulted from the operation of the mechanism suggested in the problem formulation, the possible operation of

TABLE 1

Proportion of *S*s Going Initially toward Shock Compartment or Neutral Compartment on Test

Group	*N*	Toward SC	Toward NC	*t*
I	20	3	17	3.12*
II	20	6	14	1.78
III	20	6	14	1.78
IV	20	12	8	0.89
Total	80	27	53	2.89*

* Significant at the 1% level of confidence.

any external cues which, despite the experimental precautions, might have served to mediate transfer through primary stimulus generalization must be examined. It is unlikely that any such cues were perceptually operative at the time of initial exit from the compartment (i.e., after *S*'s introduction and the closure of the compartment top), but cues, if present, might have persisted briefly in stimulus-trace form. A check on this hypothesis is provided by the experimental procedure earlier described of reversing the orientation of the runway, with respect to the room, between the shock and the test trials for half of the *Ss*. In this "different" class the position of the runway was reversed and the orientation of *Ss* when introduced into the compartments was the same with respect to the room on both the sixth training trial and the test trial, but the position of the runway had shifted. Animals responding on the basis of room orientation cues would tend to leave the compartments in the actual direction of the shock compartment. The results for the "same" and "different" (latter figures in parentheses) classes, respectively, are as follows: 8 (9) of the *Ss* in Group I, 7 (7) in Group II, 5 (9) in Group III, and 6 (2) in Group IV initially went toward the neutral compartment. The rotation of the runway after the shock trial had, thus, no consistent effect on the tendency of *Ss* to avoid going in the direction of the shock compartment. This supports the assumption that orientation with respect to the room did not noticeably determine initial exit behavior.

Despite efforts to secure a uniformly illuminated runway, a theoretical possibility existed that the end compartments, because of their being closer to the illuminated end of the experimental region, might have been slightly more illuminated than the middle. A check on any such effect is provided by grouping together the data on direction of initial exit from Groups I and II and from Groups III and IV. Of the *Ss* comprising Groups I and II, 31 of 40 went initially toward the neutral compartment (i.e., toward the middle of the runway), and of the *Ss* comprising Groups III and IV, only 18 of 40 went toward the shock compartment (i.e., toward the middle of the runway). These two proportions differ significantly ($t = 3.64$, $p = .01$) which indicates that these groups of *Ss* did not respond to the shock compartment and the neutral compartment in the same manner. Thus, this second explanation of the results obtained may be rejected.

Direction of exit as a function of the number of compartment steps removed from the shock compartment.—The experimental condition corresponding to the inferred variable of functional indirectness of memory trace connection consisted in the number of compartments between the shock compartment and the compartment into which *S* was introduced. As may be seen from Table 1, exits in the direction away from the shock compartment are relatively greatest from Compartment I and decrease to a chance level at Compartment IV. This is consistent with Assumption 6. The proportion of *Ss* initially going toward the neutral compartment in Group I is significantly different from chance at the .01 level of confidence. Tests of significance of the other three groups taken individually do not yield significant differences.

The data were analyzed to ascertain the relationship between initial exit behavior and *Ss*' orientation on being placed in the compartments. During the last three trials, 78% left the compartments in the direction

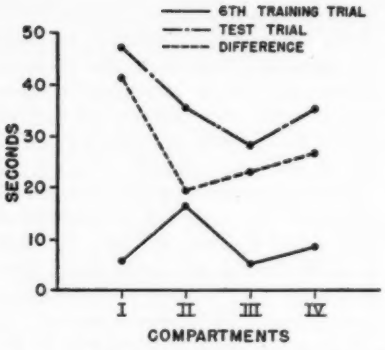

FIG. 1. Mean time spent in the compartments before initial exit

they were faced. If this tendency had operated with equal effectiveness at test, it would tend to equalize the exits in each direction. Thus, the significance of the differential directional behavior obtained is obviously enhanced.

Behavior before exit from initial compartment.—The mean time in seconds spent by *Ss* in the compartments before initial exit on both the sixth training trial and the test trial is given in Fig. 1. On the test trial, *Ss* spent more time in the introductory compartments than on the sixth training trial; perhaps this is due in part to a general depressive effect of electric shock. The lack of differentiation in latency can be attributed to primary stimulus generalization since the general structural features of the shock compartment were also components of Compartments I, II, III, and IV.

Conflict behavior, defined as approaches and withdrawals at both compartment doors and head turnings, with either or both occurring before initial exit on test, was observed with almost equal frequency in all four groups: 13 *Ss* in Group I, 15 in Group II, and 14 each in Groups III and IV. Frequency of urination and defecation was not significantly different among the four groups.

Behavior after exit from initial compartment.—The number of *Ss* of all groups reaching the shock compartment and the neutral compartment after leaving the compartments into which they were introduced is shown in Table 2. On the sixth training trial 69 of 80 *Ss* went directly to these compartments following initial exit (33 to shock compartment and 36 to neutral compartment). On the test trial only nine *Ss* went directly into the shock compartment before reversal of direction. This indicated avoidance of the shock compartment although no corresponding change is noted in *Ss'* going into the neutral compartment on this trial. This evidence indicates that many *Ss* going initially toward the shock compartment on the test trial reversed direction before receiving a direct perceptual cue of this compartment.

Distribution of behavior by compartments.—The percentages of each group of *Ss* entering the six compartments during the sixth training trial and test trial are given in Table 3. Differences in the percentages between these two trials indicate that *Ss* avoided going toward and into the shock compartment, i.e., the greatest difference oc-

TABLE 2

NUMBER OF *Ss* ENTERING SHOCK COMPARTMENT OR NEUTRAL COMPARTMENT BEFORE REVERSING DIRECTION

Group	Sixth Training Trial		Test Trial	
	Enter SC	Enter NC	Enter SC	Enter NC
I	8	8	3	13
II	11	6	3	9
III	10	8	0	10
IV	4	14	3	8
Total	33	36	9	40

curred at the shock compartment and successively less difference occurred toward the neutral compartment. Comparison of the percentages of *Ss* of each group entering the shock compartment indicates that the compartment step at which *Ss* were introduced into the runway had no apparent effect on the number entering the shock compartment or other compartments during the test trial.

The top and bottom curves in Fig. 2 represent the mean time in seconds the groups of *Ss* spent in the six compartments on the sixth training trial and the test trial respectively. The time spent by each group in the compartments into which the *Ss* were introduced is excluded. The middle curve represents the differences between these two curves. The curve of the sixth training trial shows that *Ss* tended to spend more time in the shock and neutral compartments, with the time spent in the shock compartment somewhat greater than that spent in the neutral compartment. The curve representing the differences shows the behavioral modification resulting from the shock experience.

When a curve is plotted excluding the 21 *Ss* that entered the shock compartment on the test trial, the curve is similar to that in Fig. 2, although its steepness is much increased. These *Ss* avoided this compartment on the basis of secondary processes other than a direct perceptual experience of the shock compartment.

TABLE 3

PERCENTAGES OF *Ss* ENTERING VARIOUS COMPARTMENTS

Group	*N*	SC	I	II	III	IV	NC
Sixth Training Trial							
I	20	95	*	90	80	75	70
II	20	85	90	*	85	80	75
III	20	95	100	100	*	95	95
IV	20	90	95	95	95	*	90
Mean		90	95	95	87	83	82
Test Trial							
I	20	25	*	95	85	85	85
II	20	30	50	*	85	75	65
III	20	25	35	55	*	95	80
IV	20	25	50	60	80	*	80
Mean		26	45	70	83	85	77
Diff.**		64	50	25	4	−2	5

* Introduced here.
** Difference between the averages of the 6th training trial and test trial.

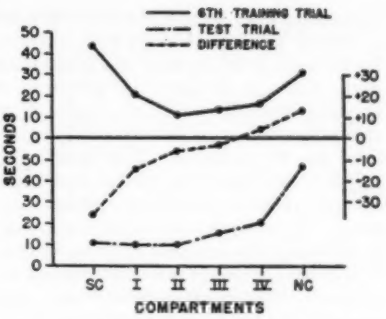

FIG. 2. Mean time spent in the various compartments after initial exit

DISCUSSION

Some of the latent-learning experiments (**8**) have suggested that behavior transfers under conditions in many ways similar to those of the present investigation. Historically, giving animals exploratory training in mazes was first used by Lashley (**5**), and later by Haney (**4**). More specifically, however, the training procedure used by both Buxton (**2**) and Seward (**7**) is principally the same as was used in the present experiment. In order to give an explanation of the behavior showing transfer (i.e., evidence of latent learning) in these studies, an evaluation of the role of both intra- and extramaze cues is necessary. An analysis of Buxton's and Seward's studies indicates that

extramaze cues were quite likely sources of stimuli which provided a basis of transfer through primary stimulus generalization. Only in Seward's investigation is concrete evidence available, however, to show that this source of stimuli was utilized by the animals. This was in his major experiment in which significant transfer effects were shown. The inadequacy of controls over extramaze cues in these studies prevents their being used as specific experiments demonstrating exclusively the transfer mechanism proposed in the problem formulation.

Experiments conducted by Miller (6) and Tolman and Gleitman (9) have yielded transfer effects mediated by the mechanism of the fractional anticipatory goal reaction. In each experiment, *S*s were trained to run to the end of a maze in order to obtain a reward. Subsequently, they were directly introduced into the goal region, shocked, and removed. An explanation of the avoidance behavior observed on a later test trial is provided by the fact that the response to shock was conditioned to the response to the earlier given reward and the former was elicited on test as a part of the fractional anticipatory goal reaction.

To summarize, none of these selected experiments entirely meets the specifications which were previously outlined in this paper, because of the possible explanation of transfer on the basis of other transfer mechanisms. No precise test of the hypothesis advanced has heretofore been made.

The specific effect of the transfer mechanism mediating avoidance behavior in the present study is significant in light of its relevance to the approach and avoidance gradients reported by Brown (1) and Dollard and Miller (3). According to Dollard and Miller, these gradients are determined by (*a*) the effect of the gradient of reinforcement, and (*b*) the effect of the gradient of primary stimulus generalization. The current study suggests, first,that the obtained avoidance behavior which was shown to have gradient properties was mediated by the functional interrelation of memory traces. This indicates that the mediation of gradient-like avoidance behavior (and probably approach behavior) is possible by memory traces, and that the effects of the mechanisms otherwise mediating these gradients need not be operating. Second, the transfer effects obtained in this study further suggest that the mechanism mediating this transfer may be operating in conjunction with the effects of the gradient of primary stimulus generalization and the gradient of reinforcement in mediating the behavior from which the approach and avoidance gradients have been derived.

Less direct mechanisms of transfer as contrasted with more direct ones (e.g., secondary stimulus generalization or the fractional anticipatory goal reaction vs. primary stimulus generalization) in mediating transfer seem to be primarily applicable to behavior which occurs in situations that are nonidentical to the original learning conditions. The present study is one which purports to demonstrate a mechanism of this type. Further detailed study is needed, however, to specify more precisely its functional properties.

Summary

The present experiment constitutes a test of the hypothesis that transfer of learned avoidance behavior may be mediated by indirect activation of memory traces under conditions which preclude explanation in terms of stimulus generalization, fractional anticipatory goal responses, or delay of reinforcement. The functional relationship between the magnitude of this transfer effect and directness of trace connection was also investigated.

Four groups of 20 rats each were given six daily nonreinforced exploration trials in a runway divided by doors into six, perceptually distinct, compartments. After training each *S* was introduced directly into one end compartment, shocked, and removed. After 15 min. it was reintroduced into one of the four middle runway compartments ($N = 20$ in each compartment). Initial direction of exit from these compartments and other behavior occurring during the 2.5-min. test period were recorded.

Statistically significant evidence of transfer of avoidance behavior as indicated by initial direction of exit was obtained. Controls indicate the inapplicability of the alternative explanatory principles mentioned above.

The hypothesis that the magnitude of the transfer effect is a positive function of the directness of connection between the compartment perceived on test and the end shock compartment (and between the corresponding memory traces) was confirmed by the differential avoidance behavior of the experimental groups.

References

1. Brown, J. S. Gradients of approach and avoidance responses and their relation to level of motivation. *J. comp. physiol. Psychol.*, 1948, **12**, 450–465.
2. Buxton, C. E. Latent learning and the goal gradient hypothesis. *Contr. psychol. Theory*, 1940, **2**, No. 2.
3. Dollard, J., & Miller, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
4. Haney, G. W. The effect of familiarity on maze performance of albino rats. *Univer. Calif. Publ. Psychol.*, 1931, **4**, 319–333.
5. Lashley, K. S. A simple maze: with data on the relation of the distribution of practice to rate of learning. *Psychobiology*, 1918, **1**, 353–367.
6. Miller, N. E. A reply to "Sign Gestalt or conditioned reflex?" *Psychol. Rev.*, 1935, **42**, 280–292.
7. Seward, J. P. An experimental analysis of latent learning. *J. exp. Psychol.*, 1949, **39**, 177–186.
8. Thistlethwaite, D. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, **48**, 97–129.
9. Tolman, E. C., & Gleitman, H. Studies in learning and motivation: I. Equal reinforcements in both end-boxes, followed by shock in one end-box. *J. exp. Psychol.*, 1949, **39**, 810–819.

(Received February 17, 1953)

Journal of Experimental Psychology
Vol. 46, No. 6, 1953

PROACTIVE INHIBITION AND ASSOCIATIVE FACILITATION AS AFFECTED BY DEGREE OF PRIOR LEARNING[1]

STEPHEN K. ATWATER

Knox College

Proactive inhibition (PI) may be defined as a decrement in the recall of a learning task as a result of the prior learning of some other task. This may be contrasted with retroactive inhibition (RI) which may be defined as a decrement in the recall of a learning task as a result of some other task being learned between the original learning and its recall.

Associative facilitation refers to a positive transfer effect of one learning task upon a second learning task, particularly where the two tasks involve paired-associate learning in which *S* must learn a new response to an old stimulus. This may be contrasted with associative inhibition in which negative transfer effects appear.

The present study is concerned with the effects of various degrees of prior learning upon the acquisition and retention of a second similar material. It was set up in much the same way as an earlier study by Underwood (**10**).

The degree of second-list learning was carried beyond one perfect trial, unlike Underwood's study, to determine whether associative inhibition changes toward facilitation with increasing degrees of prior learning when the criterion of second-list learning is one perfect trial. This was found to be the case in a study by Bruce (**2**). However, an unpublished study by G. R. Stone (**7**) seemed to indicate a reversal of this function when the degree of second-list learning reached one perfect trial.

The range chosen for the independent variable was from zero, the control condition, to a degree of overlearning which extended the scale beyond that employed by Underwood. This was done to determine whether PI shows a reversal in direction with increasing degrees of prior learning, as has been shown for RI with increasing degrees of interpolated learning (**4, 6, 8, 9**).

Procedure

Apparatus and materials.—The material to be learned was presented by means of an intermittently rotating Baker-Stone memory drum (**1**). The traditional A–B, A–C paired-associate anticipation technique was employed. The drum was timed so that the stimulus item of each pair was exposed for 2 sec., following which the stimulus and response items were exposed together for an additional 2 sec. A rest of 8 sec. followed each trial, which consisted of one complete presentation of all the paired items of a given list. Each list contained ten pairs of three-letter words, which were paired so that there was no obvious meaningful connection or formal similarity between the two words of any pair. Five different pairs of lists were constructed so that each *S* learned a new pair of lists during each of the five sessions of the experiment.

The ten pairs of words in each list were presented in three different orders to prevent serial learning and the effects of serial position which tend to decrease the amount of PI (**5**), the production of which is essential to the present study.

Subjects and experimental design.—Thirty-two undergraduate college students served as *S*s; each underwent one practice session and four experimental sessions. Each *S* served on five successive days at approximately the same hour each day.

Four degrees of prior learning were used, the degree of learning being defined by a performance criterion. Under Cond. A, the control

[1] This study was done in partial fulfillment of the requirements for the M.S. degree at the University of Oklahoma.

condition, the first list was not presented. Under Cond. B the first list was learned to the criterion of six or more correct responses on one trial. For Cond. C the criterion which had to be reached was one perfect trial plus five additional trials, and for Cond. D it was one perfect trial plus fifteen additional trials.

For all conditions the second-list learning was carried to the criterion of one perfect trial plus five additional trials, after which a 10-min. rest interval was spent in light reading. For all conditions second-list relearning was carried to the criterion of one perfect trial.

Instructions commonly given for paired-associate learning were read to each *S* on each of the five days. Special care was taken at the time of recall to insure that *S* knew which list he was expected to recall.

Conditions and lists were systematically randomized independently over the four experimental sessions. Each of the four conditions occurred once with each of the four paired lists on each of the four experimental days. This resulted in 16 different orders of presentation of the various condition-list combinations. The 32 *S*s were randomly assigned to these 16 different orders. The results were treated by analysis of variance.

Results and Discussion

Prior learning.—The mean number of trials required to reach the criterion of first-list learning under Cond. B was 6.16 (σ_m = .49); under Cond. C it was 19.78 (σ_m = 1.10); and under Cond. D it was 29.41 (σ_m = 1.00). It was expected that the materials used in this study would be learned more rapidly and retained better than those used by Underwood (10), which were paired two-syllable adjectives. Greater retention would provide a greater interlist differentiation, which, it was thought, might yield relatively less interference between the two lists, when the first list was learned to a high degree. This would increase the likelihood of obtaining a reversal in PI as has been obtained for RI; however, a comparison between Underwood's first-list learning and the present results fails to indicate that the three-letter words used here are any easier to learn than the two-syllable adjectives used by Underwood. All other relevant variables, such as length of list and rate of presentation, were the same; however, it is possible that the populations sampled by the two studies were not comparable.

Second-list learning.—Separate analyses of variance were carried out for the number of trials required to reach the criteria of 8, 9, and 10 correct responses out of 10. For these three analyses logarithmic transformations of the data were deemed necessary in order to satisfy the assumption of normality of the population of measures from which the samples were drawn. For the sake of brevity these results will not be presented here.[2] The present results significantly corroborated those of Underwood (10), which indicated an increasing associative facilitation as the degree of prior learning increased.

As stated previously, it was suspected that carrying second-list learning to the criterion of one perfect trial might produce a reversal in the effect of increasing degrees of prior learning, i.e., a reversal from increasing associative facilitation to increasing associative inhibition. This reversal did not appear.

These analyses indicated one other point which seems worthy of mention. Practice effects were very marked even between the fourth and fifth sessions (*t* significant at .01 level of confidence). This should emphasize the need for counterbalancing the effects of practice in a study of this type even though several sessions may have been spent in an attempt to minimize such effects.

Associative inhibition and facilita-

[2] Summary tables of these analyses, the means for the four conditions, the differences between means, and their probabilities, are included in the original thesis which is on file at the University of Oklahoma Library, Norman, Oklahoma.

tion in second-list learning.—If associative inhibition is present in a situation such as this, it is most likely to be found on the first few anticipation trials of the second list. Separate analyses of variance were carried out for the number of correct responses on the first and second anticipation trials. Again, the results were very similar to those obtained by Underwood (**10**), which indicated significant associative inhibition on the first two anticipation trials of the second list when prior learning was low, and significant associative facilitation on the second anticipation trial when prior learning was high. The present results showed the same trend, but were somewhat less significant.

Proactive inhibition in relearning.—An analysis of variance was carried out for the number of trials required to relearn the second list to the criterion of one perfect trial. A compound transformation of the data (log [(log X) + .005]) was necessary to satisfy the assumption of homogeneity of variance. Bartlett's test for homogeneity of variance showed this transformation to be adequate. The transformed mean for each of the conditions and the differences between these means are presented in Table 1. All the differences between the means of the various conditions were significant except those between Cond. A and B and between Cond. C and D.

TABLE 1

PROACTIVE INHIBITION IN RELEARNING

Condition	Mean*	Difference	p
		B–A .052	
A	−1.270	C–A .599	.001
B	−1.218	D–A .674	.001
C	−0.671	C–B .547	.001
D	−0.596	D–B .622	.001
		D–C .075	

* Figures represent transformed values.

Studies by Melton and Irwin (**6**) and McGeoch (**4**) have shown RI, as measured by relearning, to increase with increasing degrees of interpolated learning. In these studies the function was found to reverse when interpolated learning reached a very high degree. Thune and Underwood (**8**) were unable to find any significant RI as measured by relearning. Underwood (**10**) did not obtain significant PI for this relearning measure.

The present study produced significant PI at relearning which increased as the degree of prior learning increased. Unlike the RI studies mentioned above, no reversal of this function was obtained even when the prior list was greatly overlearned; however, PI did not show a significant increase from Cond. C to D, which might indicate a leveling of the function after high degrees of prior learning have been reached.

The fact that significant PI at relearning was obtained here, but not in the Underwood study, might be attributed to the following factors: (*a*) PI, like RI (**6, 8**), dissipates quite rapidly during relearning, as can be seen in the present study from the measure of PI at recall, which will be considered shortly; (*b*) in the present study the mean number of trials to relearn was between three and four; (*c*) it seems reasonable to assume that relearning would have taken longer in the Underwood study since the original learning of the second list was carried to a much lower degree than in the present study. Thus, in the present study, relearning was completed before the PI effects were completely lost, whereas, in the Underwood study, the PI effect in the first few relearning trials must have been masked by the later relearning trials in which PI was probably absent.

Proactive inhibition at recall.—The relationships between conditions and errors for each of the first four relearning trials are presented graphically in Fig. 1. This error measure includes failures to respond as well as incorrect overt responses. The nature of these data did not justify the use of analysis

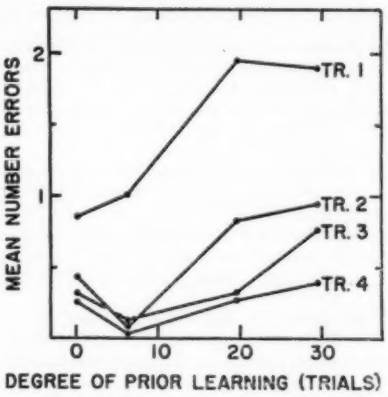

FIG. 1. Effects of the degree of prior learning upon the number of errors in recall of the second list for each of the first four relearning trials

of variance or any other parametric test; therefore, a series of nonparametric sign tests was employed to test the significance of differences between conditions. This type of test is illustrated by Dixon and Massey (**3**, pp. 247–250).

The only conditions which showed a significant PI effect were Cond. C and D on the first relearning trial (.05 level of confidence). Again, there was no significant reversal of the PI function with high degrees of prior learning, but neither was there an increase in PI from Cond. C to D. By the fourth relearning trial no differential effects due to conditions could be claimed, which is in accord with the view that PI is highly transitory.

Inspection of Fig. 1 would seem to indicate that Cond. B on the second, third, and fourth relearning trials produced a facilitation effect as compared with the control condition. The sign test did not show this effect to be statistically significant; however, the sign test is relatively inefficient since it takes into account only the direction of differences, and not their magnitude. If original second-list learning had not been carried to such a high degree, the errors on the first few relearning trials would have been greater in number and probably would have been distributed in a more nearly normal fashion. This might have justified the use of a more precise statistical technique which, in turn, might have indicated whether this "facilitation effect" was real.

Intrusions.—Underwood (**10**) has suggested that overt intrusions, or first-list responses made during relearning of the second list, should reflect the degree of interlist differentiation, which is a function of the difference between the response strengths of the two lists, and also a function of the absolute response strengths of the two lists. He says:

> It should be noted that differentiation is postulated to be low when two response systems are of near equal strength. Undoubtedly, further elaboration of this point will be required in that absolute strength will have to be considered. It is likely, for example, reasoning from Gibson's theory, that two lists having 20 trials each will be more highly differentiated than will two lists having five trials each (**10**, p. 32).

In accordance with this, the present study should yield the greatest number of intrusions under Cond. C, where the response strengths of the two lists might be presumed to be most nearly equal. Condition C did yield the greatest total of intrusions over the entire relearning period: Cond. B showed no intrusions; Cond. C produced 15; and Cond. D produced 11. This relationship did not hold for the first relearning trial. Since so few intrusions occurred, it is quite conceivable that these results may be unreliable. However, the fact that very few intrusions occurred does, in itself, support Underwood's point: (*a*) under Cond. B the first list was learned to a relatively low degree as

compared to the second list; thus, the relative response strengths should favor the second list; (*b*) under Cond. C and D both lists were learned to a relatively high degree; thus, the absolute response strengths for the two lists should be high; and (*c*) under all conditions a relatively short retention interval was employed. All of these factors, according to Underwood, should produce a relatively high degree of interlist differentiation and relatively few intrusions.

Summary

In an attempt to determine the effect of various degrees of prior learning upon associative facilitation in learning a second similar task and upon proactive inhibition in the retention of the second task, an experiment which was essentially the same as an earlier one by Underwood (10) was carried out.

The results indicated that associative facilitation in learning a second task increased as the degree of prior learning increased. Proactive inhibition increased as the degree of prior learning increased, until a relatively high degree of prior learning was reached, after which it showed no significant change. On the second, third, and fourth relearning trials the differential effects of the degrees of prior learning were not significant, which is in accord with the view that proactive inhibition is highly transitory. The results from these later relearning trials indicated, though not significantly, that facilitation may be effected by the prior learning.

References

1. Baker, T. J., & Stone, G. R. Design and construction of a research memory drum. *Proc. Okla. Acad. Sci.*, 1948, **29**, 108–113.
2. Bruce, R. W. Conditions of transfer of training. *J. exp. Psychol.*, 1933, **16**, 343–361.
3. Dixon, W. J., & Massey, F. J. *Introduction to statistical analysis.* New York: McGraw-Hill, 1951.
4. McGeoch, J. A. The influence of the degree of interpolated learning upon retroactive inhibition. *Amer. J. Psychol.*, 1932, **44**, 695–708.
5. McGeoch, J. A., & Underwood, B. J. Tests of the two-factor theory of retroactive inhibition. *J. exp. Psychol.*, 1943, **32**, 1–17.
6. Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.
7. Stone, G. R. Proactive inhibition as affected by the degree of prior learning. Unpublished manuscript, Univer. of Oklahoma, 1949.
8. Thune, L. E., & Underwood, B. J. Retroactive inhibition as a function of degree of interpolated material. *J. exp. Psychol.*, 1943, **32**, 185–201.
9. Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59**, No. 3 (Whole No. 273).
10. Underwood, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39**, 23–34.

(Received April 11, 1953)

Journal of Experimental Psychology
Vol. 46, No. 6, 1953

EFFECTS OF RESPONSE ALTERATION AND DIFFERENT INSTRUCTIONS ON PROACTIVE AND RETROACTIVE FACILITATION AND INTERFERENCE [1]

HOWARD H. McFANN

State University of Iowa[2]

This study is concerned with proactive and retroactive effects in the successive performance of motor tasks which have the same stimulus but have varying degrees of response similarity. It is also concerned with the consequences of giving specific instructions relative to the way the second task differs from the first.

Previous investigations (**3, 5**) have demonstrated significant amounts of proactive and retroactive facilitation and interference in complex motor tasks. In each case, except for a study by Dusek (**2**), the interpolated learning (IL) task has been one involving complete reversal of controls so that the responses appropriate during IL are opposite in direction to those appropriate during original learning (OL). Dusek (**2**) studied these phenomena when partial reversal of response direction was involved, but his results are not definitive.

Additional information on IL tasks not involving complete reversal of direction is supplied by Duncan (**1**). Using the star discrimeter and employing three degrees of intertask similarity, he obtained positive transfer measured by both number of correct responses and errors. Intertask similarity was defined in terms of the number of stimulus-response pairs changed from the first to the second task. It was assumed that facilitation effects would be found in this study since the task was very similar to the one used by Duncan (**1**).

On the basis of previous studies and the principle of response generalization, it was expected that the IL task involving responses opposite in direction to those appropriate in OL would lead to the greatest amount of interference. It was also expected that the tasks involving responses in IL most similar in direction to the appropriate responses in OL would result in the least interference.

Because of response tendencies acquired outside the laboratory, it was felt that self-instruction might play an important role in the performance of the tasks. Thus, the complete reversal group might not show as much interference as would otherwise be expected. Furthermore, it was assumed that under any one condition of change, the groups receiving relevant instructions as to the nature of the change would be superior in performance, at least initially, to a group receiving irrelevant instructions.

METHOD

Apparatus.—An adaptation of the Duncan (1) star discrimeter was used to define the tasks. The task requires *S* to move a vertical wobble

[1] This paper represents part of a dissertation submitted in 1952 to the Graduate School of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree. The research was directed by Dr. Don Lewis and was supported in part by the United States Air Force under Contract No. AF 33 (038)-13214, monitored by the Director of Research, Perceptual and Motor Skills Research Laboratory, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

[2] Now with the Human Resources Research Office, George Washington University.

stick into one of six possible horizontal slots (channels), the appropriate slot depending upon the color (hue) that appears in the stimulus panel. By prearrangement, any one of the six colors can be associated with any one of the six response channels. The task requires *S* to perform until a correct response is made at which time a stepping switch automatically presents a new color, requiring a movement into another slot. The response and stimulus units are mounted on a platform which can be adjusted for height for each *S*. The number of correct responses and the number of errors were obtained for each *S*.

Subjects.—Eighty male students from an elementary psychology class at the State University of Iowa were used as *S*s. They were volunteers but received two course points for each experimental session. Ten *S*s were randomly assigned to each of eight groups.

Design.—Each *S* was instructed in the manner of operation of the apparatus and told that his task was to make as many correct responses as possible. No information was given concerning the recording of errors. All trials were 20 sec. in length with 10-sec. intertrial rest intervals.

On Day 1, each of the eight groups was given 50 trials on the standard task. The standard task required *S* to learn arbitrarily chosen light-channel interconnections. A 1-min. rest period was given after 25 trials.

All groups except Group 8 participated on Day 2. Four refresher trials on the standard task were followed by a 3-min. rest period during which the controls were changed and instructions given for interpolated learning. The tasks for the IL trials were changed as follows: for Groups 2 and 5, the light-slot combinations were rotated one position (or 60°) in a clockwise direction; for Groups 3 and 6, the rotation was two positions (120°) in a clockwise direction; and for Groups 4 and 7, there was complete reversal of light-slot combination—a 180° rotation. For all these groups, the six light-slot combinations were all changed, but in a systematic fashion. In contrast, the IL task for Group 1 required *S*s to learn entirely new randomly paired light-channel combinations; none of the adjacent combinations remained as in the standard task. Prior to IL, Groups 5, 6, and 7 were given full information on how the new task differed from the old. All other groups were told that the IL task was different, but were not instructed concerning the nature of the change. Fifty IL trials were given to each group with a 1-min. rest after 25 trials.

Four refresher trials on the IL task and 50 relearning trials on the original task were given on Day 3 to all groups except Group 8. In order to determine the effect of rest on performance, Group 8 was not given any IL but had 54 RL trials on Day 3.

Results and Discussion

Comparability of groups.—To determine whether the eight groups were equivalent immediately prior to any differential treatment, a simple analysis of variance (4) was made of the number of correct responses on the last trial of OL. An F ratio of .52 ($df = 7$ and 72) was obtained which gave no basis for rejecting the null hypothesis. The means for Groups 1 through 8 were 17.2, 17.4, 17.7, 19.0, 16.2, 16.9, 17.2, and 16.7, respectively.

An additional analysis of variance made on the last 10 OL trials for the eight groups on number of correct responses yielded an F ratio of .68 ($df = 7$ and 72) also indicating that the groups were comparable.

Proactive effects in Group 1.—The assumption of equal task difficulty made possible an evaluation of proactive effects by a direct comparison of performance on the initial trials of OL and IL. It was felt that this comparison might be misleading inasmuch as the experimental groups were given four refresher trials on Day 2 prior to the beginning of IL. However, the comparison turned out to be quite appropriate. This is shown in Fig. 1 where the means of number of correct responses and of number of errors, for Group 1, on the OL and IL trials are plotted. It will be seen that the curves for correct responses are virtually parallel, with the IL curve consistently above the one for OL. The fact that the curves are essentially parallel throughout their course indicates that forgetting was of little or no importance. The difference between the two curves may be taken as a gross measure of amount of proactive facilitation.

The IL error curve, in Fig. 1, lies conspicuously above the OL error curve on the first two trials. Later, on Trials 11–29, the IL curve is consistently below the OL curve. The

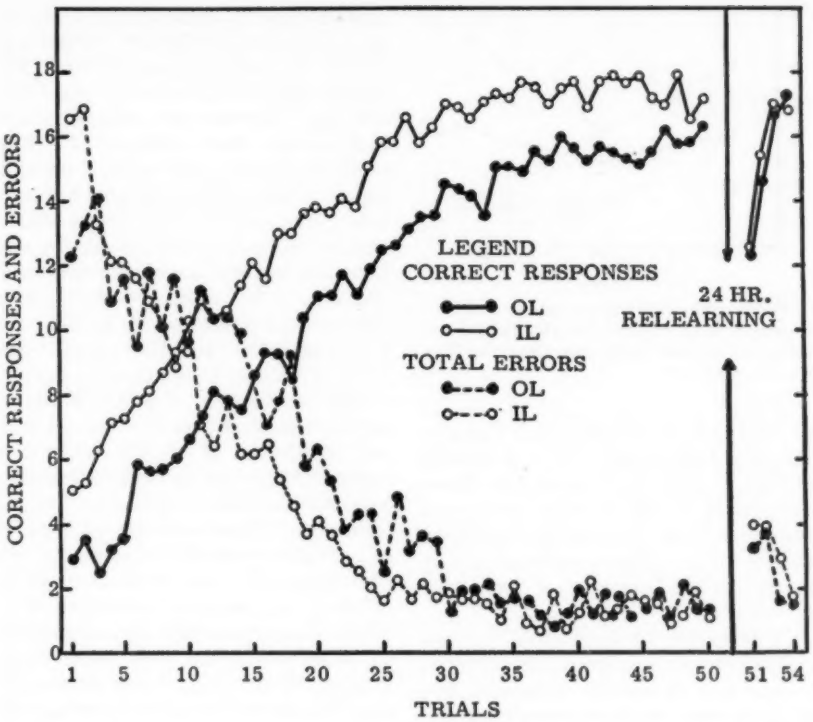

FIG. 1. The mean correct responses and errors on all OL, IL, and 24-hr. RL trials for Group 1

error curves reflect inappropriate response tendencies, that is, interference. The trend of the IL error curve is taken to indicate that proactive interference was initially high, that it dissipated rapidly, and that by Trial 11 the influence of proactive facilitation became greater than that of proactive interference. Perhaps the major point to be emphasized is that both facilitation and interference played a part in performance during IL.

The curves in Fig. 1 for Trials 51–54 show that there were marked decrements in proficiency over a rest interval of 24 hr. The number of correct responses on Trial 51 was significantly below the number on Trial 50 while the number of errors on Trial 51 was well above that on Trial 50. These differences were gone by Trial 54.

Proactive effects as related to amount of response change and instructions.— All experimental groups displayed significant proactive increments in number of correct responses when tested by the *t* test for related measures ($p \leqq .01$). As shown in Fig. 2, the groups had different amounts of gain, with Group 7 (180° relevant) showing the greatest and Group 1 (random) the least amount. The relevant instruction groups displayed greater gains than the irrelevant instruction groups. Under both conditions of instructions, the 120° groups displayed the least gain.

To determine the significance of the different increments, simple analysis

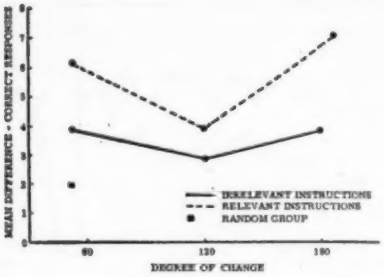

FIG. 2. Mean difference in correct responses between first IL trial and first OL trial for relevant and irrelevant instruction groups for three degrees of change

of variance was employed. The obtained F ratio of 4.39 (df = 6 and 63) indicated that there were significant differences in the number of correct responses from the first trial of OL to the first trial of IL—significant differences in the amount of proactive facilitation.

Differences between the various pairs of means were evaluated with the t test for unrelated measures, the within-groups variance estimate being used as the error term. The differences between the means for Group 7 and those for all other groups, except Group 5, were highly significant ($p \leqq .01$). There were highly significant differences between the mean for Group 5 and those for Groups 1 and 3 ($p \leqq .01$) and less dependable differences between Group 5 and Groups 2, 4, and 6 means ($.05 < p < .10$). The remaining differences could quite easily have arisen from chance ($p > .10$).

A trend analysis was made on the data for Groups 2 through 7 for the first ten IL trials. These trials were selected because an inspection of the various performance curves indicated that observable effects of different degrees of response change and of instructions were most noticeable early in IL. The analysis which is summarized in Table 1 followed Lindquist's Type III design (4, p. 281–284) especially formulated for trend data.

The first hypothesis tested was that the curves were parallel for the three different degrees of change when instructions were disregarded. The analysis yielded an F of 1.32, providing no basis for rejecting this hypothesis. The second hypothesis under test, that the general means for the three degrees of change were the same, could not be rejected ($F = 1.37$). The over-all means on IL Trials 1–10 for the 60°, 120°, and 180° groups were 10.92, 9.93, and 11.30, respectively.

The interaction (I × D) between instructions and degree of change was not significant ($F = .93$), so it was assumed that the effects of the instructions for IL Trials 1–10 were the same for the three degrees of change. An F of 2.20 gave no basis for rejecting the hypothesis that the over-all effect of different instructions was other than zero. Since the trials by instruction (T × I) interaction was highly significant ($F = 3.81$), it was concluded that the curves for the two instruction groups were not parallel.

The mean number of correct responses of the combined irrelevant instruction group was below (mean

TABLE 1

TREND ANALYSIS OF VARIANCE ON THE FIRST TEN IL TRIALS FOR MEAN NUMBER OF CORRECT RESPONSES OF GROUPS 2 THROUGH 7

Source	df	ms	F
Between *Ss*	59		
Instructions (I)	1	214.80	2.20
Degrees (D)	2	134.18	1.37
I × D	2	91.13	.93
Error	54	97.68	
Within *Ss*	540		
Trials (T)	9	142.48	45.52*
T × I	9	11.93	3.81*
T × D	18	4.14	1.32
T × I × D	18	3.36	1.07
Error	486	3.13	

* Significant at .01 level of confidence.

= 7.99) that of the combined relevant instruction group (mean = 9.90) on IL Trial 1, but by the end of ten trials was coincident with it. On IL Trial 10 the mean for the relevant instruction group was 12.77 while the mean for the irrelevant instruction group was 12.24. The difference between the Trial 1 means was tested with the *t* test for unrelated measures with the within-groups residual variance used as the error term. The obtained *t* of 8.7 indicated a highly significant difference between the means. Therefore, it was concluded that the combined relevant instruction groups on the initial IL trial were superior in performance to the combined irrelevant instruction groups.

Since the groups showed significant differences over trials, it was concluded that learning took place.

All experimental groups, except Group 2 (60° irrelevant) displayed mean proactive changes in number of errors from Trial 1 of OL to Trial 1 of IL significant at or beyond the .05 level of confidence. As seen in Fig. 3, without relevant instructions on the nature of the change in the IL task, there was some amount of proactive interference (except in Group 2 where the change was 60°); i.e., an increased number of errors at the outset of IL. The trends of the curves indicated that the interference effects dissipated rapidly.

Simple analysis of variance was used to determine the significance of the various amounts of proactive change in errors. The obtained *F* of 6.16 ($df = 6$ and 63) indicated significant differences among the mean changes. Differences between the various pairs of means were evaluated with the *t* test for unrelated measures, the within-groups variance estimate being used as the error term. Mean differences between the means of relevant and irrelevant instruction groups were all

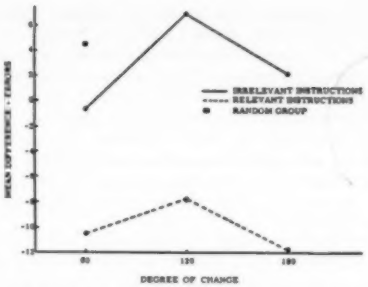

FIG. 3. Mean difference in errors between first IL trial and first OL trial for relevant and irrelevant instruction groups for three degrees of change

highly significant ($p \leqq .01$). The differences between the mean of the 60° and the means of the random and the 120° irrelevant instruction groups were also highly significant ($p \leqq .01$). The other differences could be attributed to chance factors in sampling ($p > .10$).

In summary, the presence of proactive facilitation during IL practice was taken to mean that the experimental *Ss*, during their acquisition of proficiency on the OL task, acquired not only skill in making movements in specified directions upon the appearance of different colors of lights, but also responses to general features of the stimulus situation which could be utilized during the acquisition of proficiency on the IL task. The two tasks had many features in common. Appropriate responses to these common features contributed to level of performance on both tasks.

The presence of proactive interference implied that skill in performing the OL task (that is, in executing specific directions of movement in relation to specific colors of lights) consisted, at least in part, of response tendencies which were inappropriate to the performance of the IL task.

Specific instructions on the nature of the change from the OL task to the IL task facilitated performance on the IL task. In more complex tasks, appropriate ways of responding to the "nature of the task" must usually be acquired during practice. In fact, in complex tasks, it has been found (6) that verbal pretraining is only partly effective, if effective at all, in providing *Ss* with a basis for appropriate responses in a new task.

It was expected on the basis of response generalization that the greatest amount of proactive interference would be displayed by the 180° shift group, a lesser amount by the 120° group, and

the smallest amount by the 60° group. However, the greatest amount was actually shown by the 120° group. This finding may be explained in terms of proactive facilitation: *Ss* in their past activities have often learned to make opposite responses. This view is supported by the fact that most *Ss* in the irrelevant instruction groups said, upon being quizzed at the end of the experiment, that they had expected the IL task to involve a 180° change in the light-slot combinations.

Retroactive effects.—Practice on the IL task interfered with subsequent performance on the original task. This was shown by statistically significant decrements in correct responses by all experimental groups from the last OL trial to the first RL trial. The decrements for the experimental groups were all greater than 4.3. That this was not due to passage of time was shown by the performance of a control group which received no IL training and which had a mean decrement of .3. Comparisons among experimental groups revealed that the differences were not reliable ($p > .10$).

The relevant instruction groups made markedly fewer errors on initial RL trials than did the irrelevant instruction groups. This indicated that specific instructions on the nature of the IL task were facilitative even in RL. The *Ss* probably used self-instructions which had essentially the same effects as the instructions given by *E* prior to the beginning of IL practice.

The presence of retroactive interference was taken to mean that during IL the *Ss* acquired response tendencies to the specific aspects of the IL task that were inappropriate for performance on the original task, that is, the skill developed during IL interfered with performance on the standard task during RL.

Summary

Four similar, yet different tasks, available on the star discrimeter, were used in a study of retroactive and proactive facilitation and interference in the performance of motor tasks. Eighty male *Ss* were run, 10 per group. The original (OL) and relearning (RL) task was the same for all groups. Three interpolated (IL) tasks differed in the degree to which the responses that were appropriate in the IL task differed from responses that were appropriate in the OL task. Half of the *Ss* under each degree of change received relevant instructions regarding the way the IL task differed from the OL task.

The data as a whole supported the notion that both facilitation and interference operate when practice on one psychomotor task is followed by practice on similar, yet different, psychomotor tasks. The measures used in the present study seemed to be differentially sensitive to the facilitating and interfering effects.

The results were:

1. Groups receiving irrelevant instructions prior to IL exhibited both facilitation and initial interference in IL.
2. Groups receiving relevant instructions prior to IL exhibited greater amounts of facilitation than did those receiving irrelevant instructions.
3. All experimental groups exhibited significant amounts of retroactive interference with the irrelevant instruction groups showing the greater amounts.

References

1. Duncan, C. P. Transfer in motor learning as a function of degree of first-task learning and inter-task similarity. *J. exp. Psychol.*, 1953, **45**, 1–11.
2. Dusek, E. R. Facilitation and interference in the performance of simplified tasks on the Modified Mashburn Apparatus. Unpublished doctor's dissertation, State Univer. of Iowa, 1951.
3. Lewis, D., McAllister, D. E., & Adams, J. A. Facilitation and interference in performance on the Modified Mashburn Apparatus: I. The effects of varying the amount of original learning. *J. exp. Psychol.*, 1951, **41**, 247–268.
4. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
5. McAllister, D. E., & Lewis, D. Facilitation and interference in performance on the Modified Mashburn Apparatus: II. The effects of varying the amount of interpolated learning. *J. exp. Psychol.*, 1951, **41**, 356–363.
6. Spieth, W. Effects of verbal pre-training on the subsequent learning of two mutually antagonistic motor tasks. Unpublished master's thesis, State Univer. of Iowa, 1951.

(Received March 14, 1953)

Journal of Experimental Psychology
Vol. 46, No. 6, 1953

FACTORS INFLUENCING VERBAL LEARNING FROM FILMS UNDER VARYING CONDITIONS OF AUDIENCE PARTICIPATION[1]

DONALD N. MICHAEL AND NATHAN MACCOBY

Harvard University[2] *Boston University*

Previous research has demonstrated that learning from a factual film increases with the use of audience participation procedures. Audience participation in these experiments consists of the audience viewing a film and actively practicing materials covered in the audio-visual presentation during prescribed intervals between sections of the film. When audiences are tested after being subjected to such an active review procedure, immediate retention has been shown to be higher than when an equal amount of film time is devoted to passive review (**4, 8**). Two possible explanations of this phenomenon were offered by the previous investigators. The first is that the active review procedure guarantees increased practice of the correct responses, since in the passive review procedures there is no assurance that *Ss* really practice. The second hypothesis offered was that the active review procedure increases over-all motivation or set to learn and results, therefore, in improved performance.

The present study was designed to provide data on the relative contributions to learning of motivation to learn arising from the participation situation itself compared to that of the added and emphasized practice occurring during the participation sessions. If practice is the only important factor in learning in the participation situation, then there should be no improvement in the learning of items not practiced during the participation sessions over what would occur by simply seeing the film through with no chance for practice participation. On the other hand, if motivation arising from the participation sessions is effective as a device for increasing learning, then those items not practiced should also improve since presumably such motivation would constitute a generalized participation effect, having consequences for all pertinent stimuli to which *S* is exposed.

Two other sets of variables were also investigated: (*a*) provision of knowledge of the correct response after practice (KCR) versus no such knowledge (no-KCR), and (*b*) overt versus covert practice. Further, it seemed reasonable to assume that the motivational effects of an active review situation would be more likely to be observable if other extrinsic sources of motivation were relatively low. Therefore, an attempt was made to investigate the interactions of these variables under two levels of extrinsic motivation. It was predicted, on the basis of previous research, that the expectation of a test would produce a

[1] This article reports part of a dissertation submitted by Michael to the faculty of the Department of Social Relations of Harvard University in partial fulfillment of the requirements for the Ph.D. degree. The research was directed by Maccoby. The authors wish to acknowledge their debt to Dr. Arthur A. Lumsdaine, the monitoring officer for the contract, to Dr. Fred D. Sheffield, consultant, and to the thesis advisory committee at Harvard: Prof. Gordon Allport, Dr. Eleanor Maccoby, Prof. Richard Solomon, and Prof. Samuel Stouffer. This research was supported by the United States Air Force under Contract No. AF 33(038)-22944, monitored by the Human Factors Operations Research Laboratories, Bolling Air Force Base.

[2] Now at Boston University.

higher level of extrinsic motivation than would no such expectation. By studying the interactions of these variables, in addition to their separate contributions, considerable information could be obtained on the dynamics of the active participation process.

On the basis of previous related research and theory (**1, 2, 5, 7**) the provision of the correct response following participation was expected to enhance the learning benefits of participation, primarily by providing opportunity for the correction of wrong answers and/or further opportunity for covert rehearsal of the *correct* response. The participation situation was varied by providing practice participation followed by KCR contrasted to participation without KCR. These conditions provided additional information on the contributions of motivation arising from participation as contrasted to the contributions of practice in the presence of KCR.

In previous experiments audience participation procedures have fostered overt responding (**4, 8**). There appear to have been no pertinent experiments on the relative contributions to the efficiency of learning of written practice of meaningful material compared to practice by simply "thinking" the material. If covert participation is equally as effective as overt participation, it would often be much more practicable to employ covert participation in training film showings.

Method

General procedure.—The design provided that the three experimentally manipulated factors, overt vs. covert, KCR vs. no-KCR, and announcement of a test vs. no-announcement of a test, would be arranged so that the eight participation groups would represent all eight possible combinations of the three factors.

The relative contributions to learning of practice and motivation under participation conditions were investigated in the following way. During each participation session half of the final test items covering the *previous* section of film were specifically practiced, and half of them were not. This procedure was followed for all combinations of the other factors for all experimental treatments. The relative contributions of practice and motivation could then be determined by comparing the test scores on the practiced items and the nonpracticed items with each other and with the scores on the same items obtained by control groups who saw the film, but did not have participation sessions.

Twelve groups of *Ss* were shown a film, eight of these groups under conditions of audience participation. Each of the 12 groups contained approximately four classes of about 20 high school juniors or seniors. In some treatments the *N*'s obtained for four classes were considered too small; so additional classes were obtained. In all, a total of 1,029 *Ss* were used. Approximately 2 min. after the end of the film, they were tested on their knowledge of the factual material covered by the film. The film and test were administered to *Ss* in the rooms in their high schools usually used for visual aids purposes. The assignment of classes to different experimental treatments was done on a random basis.

The film.—The film used was an edited color version of *Pattern for Survival* (distributed by Cornell Pictures), running 13 min., 44 sec. The subject matter was civilian defense against atomic bombing.

Overt and covert participation procedures.—In each of the eight experimental groups there were four participation sessions. These occurred during three breaks in the film and at the end of the film. Participation took the form of *Ss* answering orally administered practice questions covering some of the points presented in the preceding section of the film. In the overt practice groups, *Ss* wrote their answers to the practice questions on answer sheets. In the covert groups, *Ss* were instructed just to "think" the answers. Instructions to all groups stressed that the participation periods were to help them (i.e., not to test them) learn the film content. It was emphasized that the participation material covered only part of the film content.

Knowledge of the correct response.—Half of the experimental classes were provided with KCR after answering the participation questions, and half were not. The *Ss* were informed of the correct answer immediately after they had had an opportunity to answer each question.

In the eight participation groups a total of 11 sec. of participation was provided for each question. For those groups not informed of the correct response the entire 11 sec. were available for uninterrupted practice, i.e., for attempting to answer the question. In the KCR groups 9 sec. were allotted to initial practice, and the remaining 2 sec. were available for rehearsal of the correct response.

Test.—It was announced in advance to half of the groups that there would be a test at the end of the film; no announcement of a test was given to the other half of the groups. However, the groups that were told in advance that they would be tested were not told what points the test would cover. The test contained 30 orally administered questions divided into two subsets, designated as Form 1 and Form 2, selected to be as nearly equivalent as possible with respect to difficulty of the items and distribution in the film. For half of the classes in each experimental treatment the Form 1 items were the ones practiced, with the Form 2 items comprising the nonpracticed ones. For the other half of the classes in each group the reverse was true.

Control groups.—In addition to the eight experimental groups, four control groups were employed. *Control Group 1:* in order to determine the initial level of information on the material covered by the film, four classes were tested prior to the showing of the film. This provided a base line against which to compare the teaching efficiency of the film and of the various participation methods. *Control Group 2:* this group of four classes was shown the film without any participation and was tested afterwards. Fifteen seconds of blank leader were inserted in the films at each of the points where the film was interrupted for the participation groups. *Control Group 3:* this group of four classes was the same as Group 2 except that *Ss* were told in advance that they would be tested at the end of the film on its contents.

Control Groups 2 and 3 provided a base line so that the effectiveness of any of the audience participation procedures used during the presentation of the film could be compared with that of viewing the film without the use of any participation procedure.

Control Group 4: It was possible that the nonpracticed items might be meaningfully related to the practiced items. If so, there might be some information in the answers to the practiced items which could be applied to answering the nonpracticed items. Thus, some or all of the gains in the *nonpracticed* items in the audience participation groups as compared to the nonparticipation groups might be due to transfer effects of practice rather than to increased motivation to learn.

A careful comparison did not reveal any plausible meaningful relations between one set of items and the other. Nevertheless, a fourth control group was run as a check on the possibility of gains due to transfer from practiced to nonpracticed items. In this group all practiced questions were asked at the *end* of the film rather than at intervals during the course of the film. This made it impossible for the participation procedure to motivate *Ss* to learn more, since the film had already been seen before the participation practice took place. If there was to be transference of meaningfully related material, there should have been an increase in test performance on the unpracticed material in Control Group 4 classes (compared to Control Groups 2 and 3) even with no participation-derived motivation to encourage this increased learning. However, if no gains on the nonpracticed items were to occur in this group, the likelihood of transference from the practiced items would be greatly reduced.

Group sampling.—Experimental treatments were not only administered by class units rather than to individuals, but also were assigned to pre-formed class groups, taking the classes "as they came." This means that, for purposes of estimating experimental error in group-to-group comparisons, the sampling unit was the class, not the individual. For these comparisons class means rather than individual scores were the basic units of observation for sampling, and degrees of freedom were obtained from the number of classes rather than from number of individuals. However, this group-sampling restriction does not logically apply to those analyses which compare practiced versus nonpracticed items within experimental groups; here, selective factors due to the fact of group sampling would not bias the comparison, and N could thus be based on number of *Ss* rather than number of classes.

Scoring.—The test consisted of items of information directly covered in the film. All questions were of the open-end variety which could be answered with very few words. A total of six points was possible on each item, with part scores of 2, 3, or 4 points possible on many. High reliability of scoring was insured by a preliminary procedure involving the independent practice-scoring of the same sample of papers by each of the several scorers. Differences were located and discussed. This process was then repeated on a new sample until consistent scoring standards were attained. After final scoring began, a random sample of one in ten papers was independently rescored. The ratio of agreements to total possible agreements was computed in this way for each item. These ratios ranged from 91% to 100%, with a mean of 96% for the whole test.

The maximum possible score on the whole test was 180; the maximum possible score was 90 on the 15 practiced items, and 90 on the 15 nonpracticed items. Control Groups 1, 2, and 3 were tested on the same 30 items (without practice). In order to compare these groups with the experimental groups prior to calculation of final percentage scores, it was necessary to convert the scores of Control Groups 1, 2, and 3 to a base of 90 points. (This was necessary because all members of the other experimental groups practiced on 15 items, leaving the remaining 15

as nonpracticed items.) Therefore, the mean scores for classes in Groups 1, 2, and 3 were determined by separating their total scores (composed of Form 1 and Form 2 items) into subscores for Forms 1 and 2, finding the mean of each form score, and then taking the mean of the two means.

Results

There appear to be optimum conditions for efficient experimental measurement, i.e., minimum conditions for the occurrence of either ceiling or floor effects or for clustering of score averages because of too limited a range.

The lowest mean score for any of the groups is that of Control Group 1, which is 25.4% of the possible score. The highest mean score attained by an experimental group was by the overt practice with KCR group on the practiced items, a score of 77.7% of the possible score. The nonparticipation film classes in Control Group 2 had a mean score of 50.9%.

Audience participation.—The participation procedures used in this study produced a statistically significant gain ($p < .02$) in learning of 7.3% over the learning level of 52.5% (Control Groups 2 and 3 combined) attained by simply viewing the film. (Unless otherwise indicated, all statistical tests are t tests based on the differences between the means of mean scores of the classes. All p values are two-tailed.)

However, improvement in learning occurs *only* for those items that were actually practiced during the participation sessions; the items not specifically practiced did not appear to be learned any better than would occur simply by viewing the film alone.

Test announced versus test not announced.—The combined mean score for the test-announced groups was 60.0%, and for the no-test-announced groups it was 59.7%. Further, there were no significant interactions between the "test" variable and any of the other variables. Therefore, unless otherwise indicated, comparisons referred to in the text are based on the combined test-announced and test-not-announced scores for a given variable.

Knowledge of the correct response.—There was a statistically significant ($p < .01$) average gain of 16.3 percentage points from the condition of no-KCR to the condition of KCR for the practice questions (see Table 1). KCR thus appears to be a sizable contributor to learning over and above the existence of the basic kind of participation utilized in this experiment.

Even without KCR the participation procedure was effective in improving performance on the practiced items. The improvement of 7.7% for the combined no-KCR groups over the 52.5% learning level for the combined no-participation groups is significant at the .02 level.

An exposition of the data relating participation period scores and test scores is pertinent here since an explanation for this 7.7% improvement, to be discussed presently, depends on these data.

Table 2 gives the mean scores for each practice period for those groups who practiced by writing their response with and without KCR. Table 2 also gives the corresponding final test scores for the same groups. The final test was broken into the four sets of items corresponding to the four sets of items used during the participation sessions and the means of these four subsets were computed. The same was done to the test scores for the control groups who saw the film with no participation sessions.

All test scores for the KCR groups are significantly higher ($p = .05$) than the corresponding test scores for the no-KCR groups. All test scores for overt KCR and overt no-KCR groups are significantly higher ($p = .05$) than

TABLE 1

TEST SCORES FOR SINGLE VARIABLES AND IMPORTANT COMBINATIONS OF VARIABLES

Groups	Test Announced							Test Not Announced						
	N (Classes)	Practiced Items		Nonpracticed Items		All Items		*N* (Classes)	Practiced Items		Nonpracticed Items		All Items	
		Mean	%	Mean	%	Mean	%		Mean	%	Mean	%	Mean	%
No film	4	—	—	22.9	25.4	45.8	25.4							
Film	4	—	—	48.7	54.1	97.4	54.1	4	—	—	45.8	50.9	91.6	50.9
Overt KCR	5	70.8	78.7	45.1	50.1	115.9	64.3	4	70.9	78.8	47.4	52.7	118.3	65.7
Overt no-KCR	5	55.2	61.3	47.6	52.9	107.8	57.1	4	57.0	63.4	47.8	53.1	104.8	58.7
Covert KCR	4	68.4	76.0	41.4	46.0	109.8	60.9	4	65.0	72.2	46.0	51.2	111.0	61.6
Covert no-KCR	4	53.9	59.9	48.5	53.9	102.4	56.8	4	50.7	56.3	45.6	50.7	96.3	53.4
Control group 4	2	74.6	82.9	49.0	54.4	123.6	68.7							
All practice groups	18	62.2	69.1	45.8	50.9	108.0	60.0	16	60.8	67.6	46.7	51.9	107.5	59.7
All overt practice	10	63.0	70.0	46.4	51.5	107.4	59.7	8	63.9	71.0	47.5	52.8	111.4	61.9
All covert practice	9	61.1	67.9	45.0	50.0	106.1	58.9	7	57.8	64.2	45.8	50.9	103.6	57.6
All KCR	9	69.8	77.5	43.5	48.3	113.3	62.9	8	68.0	75.5	46.7	51.9	114.7	63.7
All no-KCR	9	54.6	60.7	48.1	53.4	102.7	57.1	8	53.8	59.8	46.7	51.9	100.5	55.8
All tests	18	62.2	69.1	45.7	50.8	107.9	59.9	16	60.8	67.6	46.7	51.9	53.7	59.7

TABLE 2

MEAN SCORES FOR OVERT PRACTICE PERIODS AND CORRESPONDING SCORES ON FINAL TEST

Period	Overt Practice		Combined Control Groups
	KCR	No-KCR	
Practice			
1	14.31	13.31	—
2	10.03	8.59	—
3	12.57	12.59	—
4	18.90	20.06	—
Test			
1	17.89	13.46	10.37
2	9.91	8.02	6.14
3	15.84	13.24	11.62
4	27.05	21.30	19.07

the corresponding test scores for Control Groups 2 and 3.

There is no significant difference between the practice scores obtained during the practice periods for the no-KCR groups and those of the groups receiving KCR.

Except for the practice scores for Periods 1 and 2, the practice scores for both the KCR and no-KCR groups are not significantly higher ($p = .1$) than the final scores for the straight film showing control groups.

The test scores are significantly higher ($p = .05$) than the practice scores for Periods 3 and 4 for the no-KCR groups and for Periods 1, 3, and 4 for the KCR groups. The gain from practice period to test is considerably larger in the case of the KCR groups than for the no-KCR groups.

Overt (written) versus covert ("mental") practice.—There is no significant difference between the amount learned by overt practice and that learned by covert practice ($.3 > p > .2$). For all contingencies the improvements of either overt or covert participation procedures over the no-participation control groups are significant at the .01 level.

Interaction of overt-covert participation and knowledge of the correct response.—Reference to Table 1 shows that KCR is substantially equally effective with either the overt or covert participation procedures. Further, whichever combination of these variables is considered, their effects on the nonpracticed items are substantially the same, i.e., they do not improve the learning of the nonpracticed items. Their effect on the practiced items depends primarily on the presence or absence of KCR rather than on whether the practice is overt or covert.

DISCUSSION

The results obtained in this experiment extend the findings of previous research in this area (4, ch. 9) with respect to the efficacy of practice provided by audience participation in learning simple and complex verbal material. There is good reason, then, to propose that, at least in situations where verbal material is to be learned, audience participation can be used as a means of increasing the amount of learning over and above that which would be learned when no such participation is utilized.

Since better test performance by the experimental groups (as compared to the control groups) did not occur on those items which were not practiced during the participation sessions, it is concluded that the use of the participation procedures did not produce the kind of motivation which results in increased learning of film material not specifically practiced.

At the present time the best generalization from available evidence, including both the data of the present experiment and that of other related experiments, would appear to be as follows: When participation procedures of the sort described in this research are employed with "discrete" responses, such as items of factual information under generally high conditions of motivation, the improved learning derives primarily from the conditions of practice during participation and not from motivation to learn generated by

the participation situation. The effects are confined to the material specifically practiced. However, where the material practiced in the participation sessions consists of examples of a principle or method which is to be learned, carry-over to similar, nonpracticed material can be expected (**6, 10**). Whether such motivational carry-over occurs on discrete factual material under conditions of generally low motivation has not been tested. It is reasonable to assume that the subject matter and mode of presentation in this film generated a high level of motivation to learn for most of the *S*s. If this was the case, then additional incentives to learn of the sort provided in this experiment (e.g., the test announcement) may have been incapable of adding a significant motivational increment.

The most important contributor to effective participation, under the conditions of participation studied, is KCR. When no such knowledge is available, the improvement in learning is considerably reduced.

Here, it seems, is a highly general finding: the proposition that superior learning will occur for either manual or verbal activities when the learner is given the opportunity to rectify an error by knowing that an error has been made, what the error was, and what the correct behavior should be, has not only the empirical support of this experiment, but considerable support from the various theories of learning and previous experimentation in other media as well as films (**5**).

However, improvements in learning, while not nearly as great as with KCR, occur without KCR. These improvements can be explained by the operation of the recency effect as follows: Those *S*s who practiced the *right* answer during the practice periods are more likely to know the right answer on the final test than if they had not practiced it, or if no practice period was available. There is greater likelihood that the right answer will be remembered by *more* people at the time of the practice period than would remember it if not exposed to the question again until the test period. Therefore, the practice participation periods which do not have KCR are of value because they can at least take advantage of recency.[3]

In the present experiment, overt practice does not lead to appreciably better learning than does covert practice. Further, the combination of either covert or overt practice with KCR accounted for substantial, and approximately equal, increases in learning and proved to yield the most effective participation conditions employed in this study.

While there is no supporting evidence available in the covert practice case, there is no reason to assume that, up to the time of KCR, the percentage of *S*s knowing the right answer is any different for those who practice covertly than for those who practice overtly. Since there is no significant difference between overt no-KCR and covert no-KCR, both methods appear equally efficacious in fixating the correct response previous to KCR.

Whatever increments of learning occur *after* KCR would be the result of the same kind of practice in both the overt and covert groups; namely, 2 sec. of covert practice before the next question is asked. Therefore, the first 9 sec. of the practice trial could be expected to influence learning only during the last 2 sec. of practice if there were some interaction between the practice during the last 2 sec. after KCR and the kind of practice, overt or covert, occurring during the first 9 sec. before KCR. The results indicate no such interaction since the final scores for the two forms of practice are not significantly different. Thus, it can probably be concluded that the chief importance of the practice period, overt or covert, *before* the KCR is twofold: (*a*) to permit practice which will

[3] More elaborate interscore and intrascore comparisons than those presented in this paper indicate that most of the improvement in learning for the no-KCR groups can logically be accounted for by the operation of the recency effect. However, small but evidently real learning increases occur on items practiced in Practice Periods 3 and 4 that cannot logically be accounted for by the operation of the recency effect. These increases have been ascribed to artifacts in the design, and not to the participation sessions as such.

help fixate the correct answer for those already able to give the correct response at that time, and (*b*) to get *S* sufficiently motivated to learn the response so that when the KCR is given, he will practice it during the succeeding 2 sec. of covert participation.

The comparative results for overt and covert practice seem especially important from the standpoint of film design and utilization. When overt practice procedures are employed, additional practice materials and special conditions and assistance for proper utilization of these materials are necessary. It would appear, therefore, that all the conditions for substantially increasing learning could presumably be incorporated directly into the film at fairly little cost, and no special classroom procedure or materials would be necessary. Further experimentation is necessary, however, to confirm the expectation that provisions for covert practice built into the film itself would be effective as compared to the "live" conditions utilized in this experiment.

Summary

Previous research has shown that more learning from films occurs when the viewer takes part in an "audience participation" procedure than when he does not. This improvement might result from: (*a*) increased practice during the participation periods, and/or (*b*) increased motivation to learn arising from the participation situation. This experiment was designed to assess the relative contributions of the practice and motivation factors.

Participation consisted of practice periods alternated with sections of the film, and required the audience to answer questions on the *previous* section of the film. Some *S*s wrote their answers (overt practice) and some "thought" them (covert practice). Half of the classes were told the correct answers after they had tried to answer the questions themselves, and half were not. Half the classes were told they would be tested; half were not.

Control groups were *S*s who were given the test questions but not the film, *S*s who saw the film without participation, and *S*s who had all the participation after the film.

Audience participation procedures resulted in considerable improvements in learning verbal material as compared to simply viewing the film. These increases in learning seem to be due primarily to the effects of practice and not to the effects of increases in motivation to learn.

Regardless of the other variables employed in this study, the most important factor influencing the amount of learning in this experiment was that of the provision of knowledge of the correct response (KCR) after part of the participation session. There was no significant difference between the level of learning achieved with overt practice compared to covert practice. An unsuccessful attempt was made to provide two levels of extrinsic motivation by informing some *S*s of a test to follow the film, and not informing others.

References

1. Arps, G. F. Work with knowledge of results versus work without knowledge of results. *Psychol. Monogr.*, 1920, **28**, No. 3 (Whole No. 125).
2. Carr, H. A. Teaching and learning. *J. gen. Psychol.*, 1930, **37**, 182–219.
3. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century-Crofts, 1948.
4. Hovland, C. I., Lumsdaine, A. A., & Sheffield, F. D. *Experiments on mass communication.* Princeton: Princeton Univer. Press, 1949.
5. Hovland, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 613–689.
6. Jaspen, N. Effects on training of experimental film variables. Study II: Verbalization, "how-it-works," nomenclature, audience participation, and succinct treatment. SDC Technical Report No. 269-7-11, Office of Naval Research Project designation NR-781-005. Pennsylvania State Coll., 1950.
7. Judd, C. H. Practice without knowledge of results. *Psychol. Monogr.*, 1905, **7**, No. 1 (Whole No. 29), 185–198.
8. Lumsdaine, A. A., & May, M. A. Do motivation and participation questions increase learning? *Educ. Screen*, 1947, **26**, 256–259, 274–283.
9. McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.
10. Roshal, S. M. Effects of learner representation in film-mediated perceptual-motor learning. SDC Technical Report No. 269-7-5, Office of Naval Research Project designation NR-781-005. Pennsylvania State Coll., 1949.
11. Wiener, N. *Cybernetics.* New York: Wiley, 1948.

(Received March 16, 1953)

Journal of Experimental Psychology
Vol. 46, No. 6, 1953

TRANSFER AS A FUNCTION OF TYPE AND AMOUNT OF PRELIMINARY EXPERIENCE WITH TASK STIMULI[1]

ALBERT E. GOSS

University of Massachusetts

Miller and Dollard's (**2, 13**) statement of the mechanism of the acquired distinctiveness of cues leads to the suggestion that the acquisition of discriminative verbal responses to similar external stimuli should, by reducing the similarity of external plus verbal response-produced stimulus patterns, facilitate the subsequent learning of other discriminative responses to those same environmental cues.

Experimental findings which directly or indirectly suggest that experimentally acquired discriminative verbal responses facilitate the acquisition of discriminative motor responses have been reported by Pyles (**15**), Gagné and Baker (**4**), Baker and Wylie (**1**), and Rossman and Goss (**16**). Also corroborative was Wickens and Briggs' (**19**) observation that verbalization may decrease generalization in post-conditioning tests of the efficacy of sensory preconditioning.

Although Thompson (**18**) found that learning names for puzzle pieces did not result in significant positive transfer to a puzzle-assembly task, it is probable that the discriminative verbal responses were not sufficiently strong to influence the course of motor learning. Similarly, the negative findings of Lawshe and Cary (**12**) and the inconclusive results of Smith and Goss (**17**) may have been due to failure to establish stable verbal discriminations.

It is conceivable, however, with the Pyles study as a possible exception, that instead of, or in addition to, distinctiveness based on experimentally acquired verbal discriminations the apparent facilitative effects of premotor experiences may have reflected the influence of either one or both of two other processes which were neither eliminated nor introduced explicitly. One of these processes can be identified as acquired distinctiveness based on the premotor activation of pre-experimentally acquired discriminative verbal responses. Warm-up in the form of postural adjustments and/or receptor-orienting responses occasioned by exposure to the external stimuli during premotor experiences is the second of these processes (**11**).[2]

In view of the possible explanatory significance of prior habit arousal and/or warm-up, it seemed desirable to ascertain whether experimentally introduced discriminative verbal responses resulted in positive transfer to a discriminative motor task over and above facilitative effects attributable to one or both of these additional processes. Further, if operative and facilitative, there appeared to be the possibility that—particularly when new verbal discriminations were relatively weak—prior habit arousal or warm-up might inhibit newly learned verbal discriminations. Accordingly, in order both to minimize this possibility and

[1] This research was supported in part by a grant from the Research Council of the University of Massachusetts. The writer wishes to express his indebtedness to Mr. James Martin for his aid in carrying out this study and to Mr. Joseph Mach who was responsible for the construction of the apparatus.

[2] The possible explanatory significance of warm-up processes was hinted at by Rossman and Goss (**15**). In addition, suggestions by Drs. Robert M. Gagné and Neal E. Miller (personal communications) were of great value.

at the same time to explore the relationship between strength of newly acquired responses and amount of facilitation, an experimental design which called for three levels of mastery of verbal discriminations and corresponding degrees of prior habit arousal and warm-up experiences was employed.

Method

Subjects.—The *Ss* were 104 students drawn largely from classes in introductory psychology at the University of Massachusetts. All *Ss* were naive with respect to the general purpose of the study and to specific procedural details. Because of apparatus breakdowns and apparent failure to understand instructions, four *Ss* were discarded.

Apparatus and stimulus materials.—The experimental apparatus (Fig. 1) was a modification of the verbal-motor discrimination device described by Gagné and Baker (4). When a solenoid-activated shutter was opened, light stimuli for both paired-associate verbal learning and for motor learning were presented through a translucent plastic window in the lower circular aperture. Paired nonsense-syllable stimuli for verbal responses were exposed in the upper aperture by means of a second solenoid-activated shutter.

During verbal learning an electrically driven cam arrangement initiated successive sequences which consisted of a 4-sec. presentation of the light stimulus, during the last 2 sec. of which the syllable was exposed, followed by a 4-sec. interstimulus interval.

The length of each motor learning trial was determined by *S*'s response time with the consequence that interstimulus intervals were not automatically controlled and thus varied from 4 to 6 sec. At the beginning of each motor learning trial *S*'s thumb was pressed against the push-button and the rest of the hand lay on the toggle-switch panel roughly equidistant from the switches.

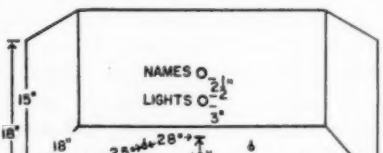

Fig. 1. Diagram of *S*'s side of the modified Gagné-Baker verbal-motor discrimination apparatus

Located behind the stimulus panel and hence not shown in the diagram were devices for the selection of light and syllable stimuli, lights which indicated thumb pressure on the push-button and toggle-switch selection, and a Standard Electric timer.

The *Ss* were seated about 12 in. in front of the apparatus which was on a small table in the approximate center of a lightproof room. There was one 100-w. overhead lamp.

Polaroid and translucent plastic filters were used to vary the intensity of light transmitted from a 60-w. bulb and thus to produce the four similar but discriminable intensities of white light which served as the stimuli for both verbal and motor responses. A point about 17 in. in front of and 11 in. above the light source was selected as representative of the location of the average *S*'s eyes. At that point the best estimates of intensity were 1.5, 1.9, 2.5, and 3.4 ft.-candles for the four brightnesses. Successive differences between logs of these intensities were approximately equal. At the same position the overhead light alone produced an intensity of approximately 1.2 ft.-candles.

Cuf, jer, nig, and *vol,* the nonsense-syllable stimuli, had Glaze (7) association values of 80% and were assigned to the 1.9, 3.4, 1.5, and 2.5 ft.-candle lights, respectively.

Each of the four lights, with or without its nonsense-syllable associate, was exposed once in each block of four trials. The order of associates or lights within each block was one of the 24 permutations of the order of four events. A single exposure of one of the paired associates or of a light alone was considered a trial.

Procedure for premotor experiences.—The design of the experiment is summarized in Table 1. The verbal learning groups were required to learn to discriminate among the four similar light intensities by means of paired nonsense-syllable responses. Group VL-9 learned to a criterion of at least 9 of 12 correct anticipations over three successive four-trial blocks; Group VL-11 learned to a criterion of at least 11 of 12 correct anticipations over three successive four-trial blocks; and the 100% overlearning Group VL-0 was given twice the number of trials to the 11/12 criterion plus the 12 criterion trials.

The instructions and experimental conditions for the seeing-and-discriminating groups were designed to activate previously learned responses for designating stimulus differences but without reinforcement of those responses. Specifically, these *Ss* were told that they were participating in an experiment to determine how well they could judge stimulus differences, and therefore that they were to pay close attention to the series of

TABLE 1

SUMMARY OF TYPE AND DEGREE OF PREMOTOR LEARNING

Group	Type of Premotor Experiences	Degree of Premotor Learning	Motor Learning Trials
VL-9	Verbal learning	9/12 Criterion	48
VL-11	Verbal learning	11/12 Criterion	48
VL-0	Verbal learning	100% Overlearning	48
SD-9	Seeing-and-discriminating	Same as VL-9	48
SD-11	Seeing-and-discriminating	Same as VL-11	48
SD-0	Seeing-and-discriminating	Same as VL-0	48
S-9	Seeing	Same as VL-9	48
S-11	Seeing	Same as VL-11	48
S-0	Seeing	Same as VL-0	48
C	None		48

stimuli to be presented in order to state not only how the stimuli differed but also, and more importantly, how many different stimuli had been exposed. Thus instructed, the 9 of 12 (SD-9), 11 of 12 (SD-11), and overlearning (SD-0) seeing-and-discriminating groups were presented with the lights alone for the same mean number of trials as was administered to VL-9, VL-11, and VL-0, respectively.

Instructions and conditions for *S*s in the seeing groups were designed to arouse postural adjustment and receptor exposure responses. Thus they were told that the purpose of the experiment was to investigate the effects of seeing stimuli, and hence that they were to pay close attention to each one of the series of to-be-presented lights. The same average number of trials were given to the 9 of 12 (S-9), 11 of 12 (S-11), and overlearning (S-0) seeing groups as were required for VL-9, VL-11, and VL-0, respectively.

Procedure for motor learning.—Immediately after *S*s of these groups had completed their premotor learning experiences, they were introduced to the discriminative motor task by means of appropriate instructions. The instructions required approximately 1 min. The discriminative motor responses were the selections of different toggle switches for each intensity. The *S*s were instructed that the correct switch for a given intensity was the one which turned off the light. Switch 1 turned off the next brightest light, Switch 2 the darkest, Switch 3 the next darkest, and Switch 4 was correct for the brightest stimulus. The correction technique was employed by permitting *S* to continue selection responses on each trial until the switch turning off the light had been chosen.

The control (C) group which had had no prior experiences with the lights was given the same instructions.

The 48 motor learning trials were made up of 12 successive four-trial blocks with each block consisting of a different order of presentation of each of the four intensities.

Assignment of subjects.—With the proportion of males (six to each group) and females controlled, *S*s were randomly assigned to one of the ten groups. They were then run in ten consecutive cycles with a cycle defined as the performances of one *S* in each of the ten conditions. Within each cycle each *S* in SD-9 and S-9, in SD-11 and S-11, and in SD-0 and S-0 was given the same number of trials as had been required for the randomly matched *S* in the corresponding verbal learning group to reach the appropriate criterion.

RESULTS

Verbal learning.—The mean number of trials to criterion and mean number of total trials for the verbal learning groups are summarized in Table 2. While not given additional or overlearning trials at the mastery level, *S*s of VL-11 had only slightly less mean total experiences than those of VL-0. This somewhat unanticipated outcome was attributable to the relatively slow learning of three *S*s who required over 100 trials to reach criterion. Since *S*s in the SD-11 and S-11 conditions were given the same number of trials as VL-11 *S*s, the 80.0 mean light experiences of those groups were also slightly less than 84.4 mean experiences of the SD-0 and S-0 groups. The *S*s in

TABLE 2

MEANS AND *SD*'S OF TRIALS TO VERBAL LEARNING CRITERIA AND FOR TOTAL VERBAL LEARNING TRIALS

Group	Trials to Criterion		Total Trials	
	Mean	*SD*	Mean	*SD*
VL-9*	34.4	17.0	47.2	17.6
VL-11	68.0	35.4	80.0	35.4
VL-0*	36.8	10.1	84.4	20.1

* Because two *S*s of VL-9 were each given four additional trials, total trials are slightly greater than trials to criterion plus 12 and the *SD*'s for the two measures differ. A somewhat similar error in trial administration for one *S* of VL-0, which was not discovered until after the completion of the study, accounts for the slight discrepancies in means and *SD*'s of trials to criterion and total trials for that group.

SD-9 and S-9 had 47.2 mean presentations of the light stimuli.

Two further aspects of the verbal learning processes should be noted. The first concerns the relationships between number of verbal learning trials and some frequency-derived measure of both correct and generalized responses. Generalized responses were defined as nonsense-syllable responses elicited by a given intensity which, while not the correct response to that intensity, were the paired-associate responses to the other three intensities. To ascertain this relationship Hilgard and Campbell's (**8**) modification of Vincent's procedure was used to compute percentages of correct and generalized responses for successive tenths of trials to the 11/12 criterion (Fig. 2) for VL-11 and VL-0.[3] Examination of Fig. 2 indicates that percentages of generalized responses increased to a maximum at the third tenth and tended to decrease irregularly from that point. With some inversions the correct responses appear to have a positive monotonic relationship to tenths of trials.

The second aspect is the relationship of generalized responses to each intensity to the similarity between the eliciting intensity and the stimulus intensities for which the generalized responses were correct. To illustrate this relationship, generalized responses to each intensity, expressed as percentages of total correct responses to that intensity, have been plotted in Fig. 3. Because four trials were necessary for *S*s to make each syllable response at least once, generalized responses during these trials were not included. Also, in order to depict the influence of learning, separate curves have been plotted for the first and second halves of all but the first four trials to the 11/12 criterion. The curves in the lower right of Fig. 3 represent tendencies for the darkest (D) stimulus to evoke the syllable responses which were correct for successively brighter stimuli. Generalized responses for stimuli successively removed from the next darkest (ND), next brightest (NB), and brightest stimuli (B) have also been plotted.

For each intensity, as physical similarity between that intensity and the other stimuli decreased, there was a

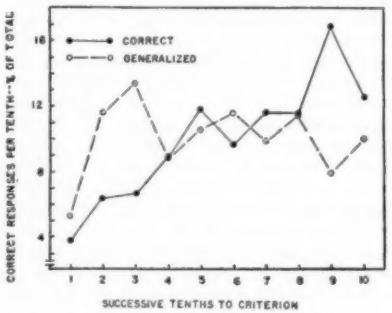

FIG. 2. Percentages of correct and generalized responses for successive tenths of trials to criterion

[3] Since VL-11 and VL-0 had the same learning criterion, data for both groups were combined. One *S* was eliminated because of difficulties in the interpretation of the scoring of several responses.

decreasing tendency for the stimulus presented to elicit responses to the other stimuli. Assuming equality of differences between successive abscissa points the gradients appeared to be negatively accelerated. In general, second-half gradients were steeper than those for the first half. Both first- and second-half trends for the intermediate intensities were less steep than corresponding portions of the gradients for the two extreme intensities. Thus, in absolute numbers, for the two halves combined, fewer correct responses (173 vs. 252) and more generalized responses (155 vs. 109) were made to the intermediate intensities.

Motor learning.—Table 3 summarizes means of errorless trials and errors for the entire 48 motor learning trials for each of the ten treatments. An errorless trial was one in which no incorrect switch choices were made, i.e., the first choice was the correct choice. Errors were the selections of incorrect switches, and up to three errors could be scored on a given trial. Because the SD-11 and SD-0 and the S-11 and S-0 pairs had essentially equivalent amounts of premotor experience, combined means for these pairs have also been computed. In Fig. 4 mean errorless trials and mean errors have been plotted as functions of level of verbal learning or amounts of seeing-and-discriminating or seeing experiences. Group VL-0 had more errorless trials and fewer errors than either the other verbal learning groups or the seeing-and-discriminating, seeing, and control conditions. All groups with premotor experiences performed at a higher level than Group C. Amount of seeing-and-discriminating or seeing experience, however, had little apparent effect on motor learning.

Differences among mean errorless trials and errors for the nine premotor experience groups were tested by means of 3 × 3 factorial analyses of variance (3). These analyses permitted the rejection of the null hypothesis with respect to differences among both errorless trial and error means for all nine groups and for the three different premotor treatments at or beyond the 1% level for appropriate *df* (Table 4). Further analysis of the

TABLE 3

ERRORLESS TRIALS, AND ERRORS FOR 48 MOTOR LEARNING TRIALS

Group	Errorless Trials		Errors	
	Mean	*SD*	Mean	*SD*
VL-9	31.0	4.4	21.2	6.9
VL-11	32.0	3.0	19.8	4.1
VL-0	36.3	4.0	14.4	6.2
SD-9	29.8	7.7	24.3	12.6
SD-11*	29.9	4.5	24.4	8.1
SD-0*	31.1	6.8	22.5	10.8
S-9	28.2	8.5	27.9	14.5
S-11†	27.9	6.8	28.5	13.7
S-0†	23.0	8.0	36.7	14.6
C	18.9	4.9	44.3	9.8

* Means and *SD*'s for combined groups SD-11 and SD-0 are 30.5 and 5.8 (errorless trials) and 23.4 and 9.6 (errors).

† Means and *SD*'s for combined groups S-11 and S-0 are 25.4 and 7.5 (errorless trials) and 32.6 and 14.8 (errors).

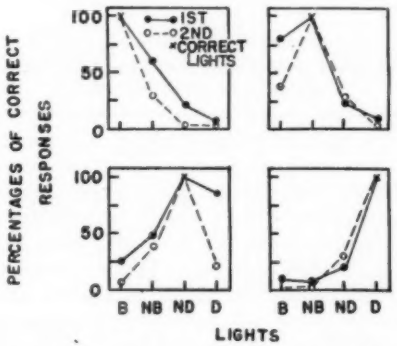

FIG. 3. Percentages of generalized verbal responses to the bright (B), next bright (NB), next dull (ND), and dull (D) intensities as a function of stimulus similarity for the first and second halves of all but the first four trials to criterion

between-groups sum of squares into sources of variation based on types of premotor experience, amounts of premotor experience, and interaction of types and amounts resulted in significant F's only for types of premotor experience.

When pairs of errorless trial and error means were then compared by one-tail t's the following results were obtained: (*a*) for both performance measures VL-0 was superior to the achievement of all other groups singly or when combined (t's from 1.95 to 8.25 with corresponding p's from .05 to <.01 for appropriate df); (*b*) with the exception of comparisons with S-0, the Ss of the control group were significantly inferior with respect to both trials and errorless trials to those who had premotor experience (t's from 2.81 to 8.25, all significant at the 1% level for appropriate df); and (*c*) differences among the remaining conditions either failed to reach the 5% level of significance, were inconsistent for errorless trial and error measures, or were inconsistent for comparisons involving single and combined groups.

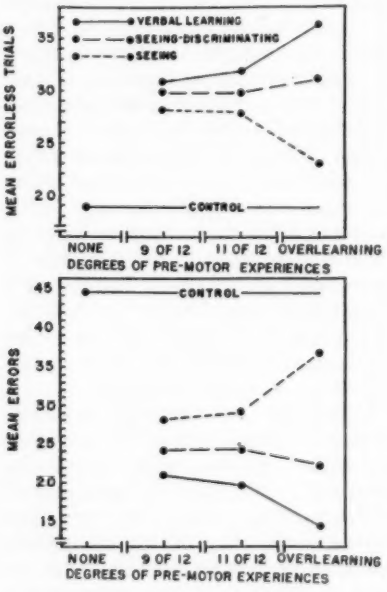

FIG. 4. Mean errorless trials and errors for 48 learning trials for each type of premotor learning experience as a function of degree of verbal learning mastery or of seeing-and-discriminating or seeing experiences

In Fig. 5 the *first* incorrect or generalized motor responses to each intensity for both halves of motor learning have been plotted as functions of stimulus similarity. Selection of an incorrect switch was considered a generalized motor response. Because type rather than amount of premotor experience was significantly related to rate of motor learning, data for each of the three groups which had different amounts of the same premotor experiences were combined to obtain the curves for the verbal learning, seeing-and-discriminating, and seeing conditions. In all curves ordinate values are generalized responses computed as percentages of correct responses.

In general, again assuming equality of abscissa differences, both first- and second-half trends for each of the combined premotor experience groups

TABLE 4

ANALYSES OF VARIANCE FOR MEAN ERRORLESS TRIALS AND MEAN ERRORS

Source	*df*	Mean Square	*F*
Errorless trials			
Groups	8	120.68	2.73*
Types	2	342.88	7.75*
Amounts	2	3.31	
Interaction	4	68.27	
Error	81	44.22	
Errors			
Groups	8	386.76	2.97*
Types	2	1194.74	9.17*
Amounts	2	.74	
Interaction	4	189.93	
Error	81	130.32	

* Significant at the 1% level.

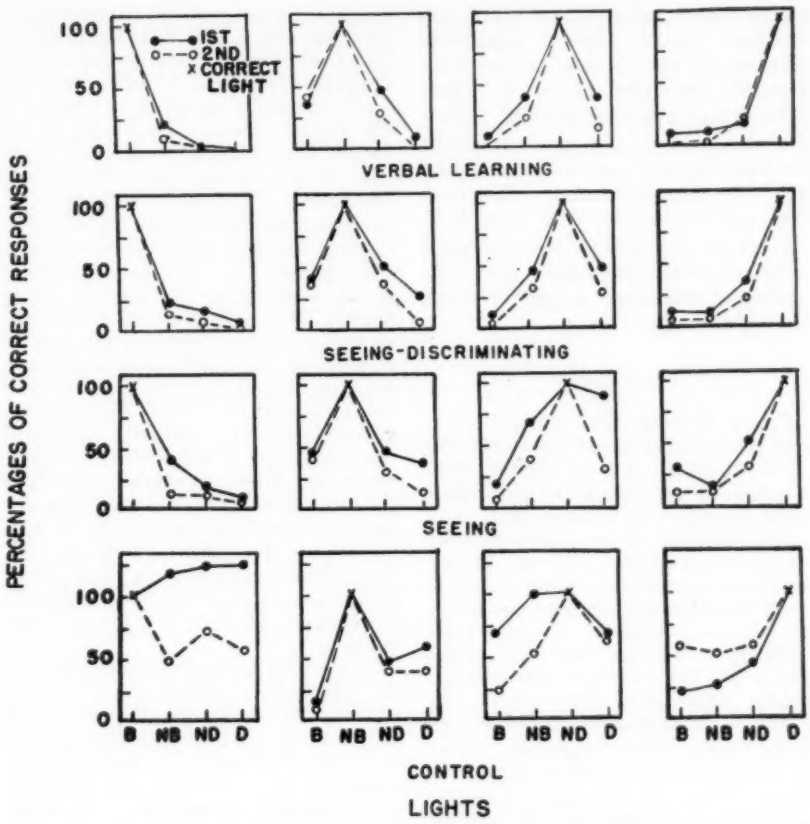

FIG. 5. Percentages of generalized motor responses to each of the four intensities as a function of stimulus similarity for each type of premotor learning experience for the first and second halves of the 48 motor learning trials

appear to be negatively accelerated functions of decreasing similarity between the eliciting stimulus and the stimuli for which the generalized responses were correct. The first- and second-half gradients about each intensity for the combined verbal learning groups appear steeper than comparable gradients for the combined seeing-and-discriminating groups, which in turn seem to have slightly higher slope values than comparable trends for the combined seeing groups. Without exception, the first-half gradients about each of the four intensities for each of the three combined groups are less steep than the second-half trends. Since they were based on fewer correct and more generalized responses, the gradients about the middle intensities are less steep than corresponding sections of the trends for extreme values.

For the control group the curves about each intensity for both halves are less steep than comparable first- and second-half curves for the pre-

motor experience groups. While second-half trends appear negatively accelerated, this is not the case for first-half gradients about the next darkest and brightest stimuli. With the exception of the gradient about the least intense light the second-half trends are somewhat steeper than corresponding ones for the first half. Unlike the curves for the premotor experience conditions, there is no suggestion of less steep gradients about the middle intensities.

Discussion

Verbal learning.—The upward-downward trend of generalized responses to the 11/12 criterion was consistent with Gibson's (5) hypothesis and with subsequent experimental findings (6).[4]

On the basis of the principle of primary stimulus generalization it would be predicted that each of the similar stimuli of a task should evoke the responses to the other stimuli. Further, the greater the similarity of two or more stimuli the greater the likelihood of each stimulus evoking responses to other stimuli. In accordance with these expectations each of the four stimulus intensities elicited the correct responses for each of the other stimuli. More significantly, the greater the similarity of intensities the greater the tendency for each intensity to evoke the responses to other intensities. It was also observed that the gradients about the extreme intensities were steeper than those about the intermediate values. A possible explanation of this finding stems from the fact that generalization of responses to the two intermediate intensities was from two adjacent stimuli one step removed and a third two steps removed in contrast to generalization to extreme stimuli of responses to intensities one, two, and three steps distant. Thus, the greater net strengths of generalized responses to intermediate stimuli than to the least and most intense lights should lead to the observed relationship of more frequent generalized responses and less frequent correct responses to intermediate values. Because knowledge of the shapes of stimulus generalization gradients about varying intensities is still uncertain, it appears premature to attempt more quantitative explanations.

Motor learning.—When compared with the performance level of the control group, warm-up in the form of seeing stimuli, the activation of pre-experimentally acquired discriminative verbal responses by means of instructions to see-and-discriminate, and the acquisition of experimentally introduced discriminative verbal responses apparently facilitated the acquisition of discriminative motor responses. However, statistical analysis indicated that only overlearning of experimentally introduced discriminative verbal responses had apparent significant positive transfer effects over and above facilitation attributable to comparable amounts of prior habit arousal or of warm-up.[5]

The failure to obtain greater facilitative effects from verbal learning to the 9 of 12 and 11 of 12 levels of mastery may have been due to the fact that the newly acquired verbal responses were not sufficiently strong to compete with the pre-experimentally established discriminations, and hence were not operative during motor learning. Also, as motor learning progressed, they may have been weakened by the dissipation of inhibitory processes (5).

Motor performance levels did not increase within the range of from 50 to 80 seeing-and-discriminating or seeing experiences. Thus, it appears likely that maximum facilitative effects attributable

[4] Recently Norris and Arnoult (13) have reported findings which suggest that the present results and Gibson's may be dependent on a procedure which permits omissions.

[5] In a check experiment a second VL-0 group proved superior to other conditions. However, Holton (9) employed essentially similar procedures and did not obtain a VL-0 motor performance level which would have been statistically significant in the present study. Accordingly, despite the statistically significant superiority of VL-0 reported here, more definite conclusions should await additional data.

to these processes are reached at or before 50 of such experiences. Because the seeing-and-discriminating and seeing treatments have such marked facilitative effects, however, their explicit introduction as controls for the assessment of the influence of experimentally acquired verbal discriminations would seem a desirable procedural refinement.

With respect to the mechanism of acquired distinctiveness, the present results seem corroborative not only because of the greater facilitation attributable to verbal overlearning, but also because it appears tenable to assume that seeing-and-discriminating arouses preexperimentally acquired discriminative verbal labels. Further, although not checked explicitly herein, it appears likely that seeing stimuli may have also activated prior discriminative habits.

If it is assumed that there was decreasing verbal involvement in the verbal learning, seeing-and-discriminating, seeing, and control conditions respectively, additional indirect support for acquired distinctiveness is provided by the finding that the gradients of generalized responses about each intensity for combined verbal learning, combined seeing-and-discriminating, combined seeing, and control treatments appeared to be of decreasing steepness.

Summary

The problem investigated was whether different degrees of acquisition of experimentally introduced discriminative names for similar light intensities would prove a significantly more effective means of facilitating the subsequent learning of discriminative motor habits than corresponding amounts of activation of pre-experimentally acquired discriminative verbal responses or of warm-up.

One hundred undergraduate students were randomly assigned to one of ten conditions. Three groups learned different nonsense-syllable responses to each of four light intensities to three levels of mastery prior to 48 motor learning trials in which discriminative switch-selection responses to lights were to be acquired. Three groups had comparable amounts of premotor experiences in seeing-and-discriminating among the intensities, and three additional groups had parallel amounts of seeing or warm-up experiences. The *Ss* in the tenth group were not exposed to the light stimuli prior to their introduction to the motor discrimination task.

All premotor experience groups learned the motor task with more errorless trials and fewer errors than the control group. The motor learning performance of the verbal learning group which had the greatest mastery of discriminative verbal responses prior to motor learning was significantly superior to the achievements of remaining premotor experience groups. In general, differences among the remaining verbal learning, the prior habit arousal, and warm-up groups were not statistically significant.

References

1. Baker, K. E., & Wylie, R. C. Transfer of verbal training to a motor task. *J. exp. Psychol.*, 1950, **40**, 632–638.
2. Dollard, J., & Miller, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
3. Edwards, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
4. Gagné, R. M., & Baker, K. E. Stimulus pre-differentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.
5. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
6. Gibson, E. J. Intra-list generalization as a factor in verbal training. *J. exp. Psychol.*, 1942, **30**, 185–200.
7. Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 517–567.
8. Hilgard, E. R., & Campbell, A. A. Vincent curves of conditioning. *J. exp. Psychol.*, 1937, **21**, 310–319.
9. Holton, R. B. The effects of five types of verbalization on acquisition of discriminative motor responses. Unpublished master's thesis, Univer. of Massachusetts, 1953.
10. Hull, C. L. *Essentials of behavior*. New Haven: Yale Univer. Press, 1951.
11. Irion, A. L. The relation of "set" to retention. *Psychol. Rev.*, 1948, **47**, 336–341.
12. Lawshe, C. H., & Cary, W. Verbalization and learning a manipulative task. *J. appl. Psychol.*, 1952, **36**, 44–46.
13. Miller, N. E. Theory and experiment relating psychoanalytic displacement to

stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

14. Norris, E. B., & Arnoult, M. D. Overt errors as a measure of generalization. *Amer. Psychologist*, 1952, **7**, 270. (Abstract)
15. Pyles, M. K. Verbalization as a factor in learning. *Child Develpm.*, 1932, **3**, 108–113.
16. Rossman, I. L., & Goss, A. E. The acquired distinctiveness of cues: the role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.*, 1951, **42**, 173–182.
17. Smith, S. L., & Goss, A. E. The role of the acquired distinctiveness of cues in the acquisition of a motor skill in children. *J. genet. Psychol.*, in press.
18. Thompson, L. The role of verbalization in learning from demonstrations. Unpublished doctor's dissertation, Yale Univer., 1944.
19. Wickens, D. D., & Briggs, E. G. Mediated stimulus generalization as a factor in sensory pre-conditioning. *J. exp. Psychol.*, 1951, **42**, 183–186.

(Received March 23, 1953)

Journal of Experimental Psychology
Vol. 46, No. 6, 1953

DISTRIBUTION OF PRACTICE PRIOR TO SOLUTION OF A VERBAL REASONING PROBLEM[1]

A. B. SHAKLEE AND B. EDSON JONES

University of Denver

Our knowledge concerning distribution of practice on tasks which require higher order thought processes rather than more strictly rote memorization is founded on a small number of studies. The investigations traditionally cited in this connection and some more recent ones have dealt with: mechanical puzzle learning, Cook (**1**), Ericksen (**4**); artificial language, Garrett (**5**); concept formation, Oseas and Underwood (**9**); and the development of negative sets through practice on a series of similar rational problems, Kendler, Greenberg, and Richman (**6**). All of these, with the exception of Oseas and Underwood, claimed an advantage for massed practice.

Although these studies are listed as a group, one cannot assume a high degree of communality in the tasks employed. The names of the tasks suggest some gross differences, and examination of the materials suggests others. The problem box used by Ericksen seems to involve a large element of chance in obtaining the first solution. With Cook's puzzles, chance factors could conceivably be important in the first solutions by at least some *S*s. Furthermore, the various tasks probably differ considerably in the extent to which negative sets influence performance.

Study of the literature suggests that there have been no experiments on different ways of distributing practice as related to the solution of a single deductive-type problem. This is not too surprising in view of the scarcity of evidence of any kind relating to deductive tasks (**7**).

It seems also that distribution of effort applied to a problem prior to its first solution has not been investigated. To do this requires a practice schedule somewhat different from those commonly used in distribution experiments. Furthermore, as Underwood (**10**) points out, published work on problem solving (**1, 4**), as differentiated from concept formation, has compared massing with only one distributed condition. This raises questions concerning the shapes of distribution curves which might be obtained where several distribution conditions were investigated.

The present study seeks to determine the effects of three degrees of distribution of practice prior to first solution of a deductive-type problem. It was decided that the problem used should conform as closely as possible with the following criteria: (*a*) be relatively free from mechanical manipulation, (*b*) be difficult to recall in its essential details, (*c*) possess sufficient variables and answers to minimize the probability of solution by chance, (*d*) minimize deliberate sets that would inhibit solution, and (*e*) be sufficiently difficult that no *S* would solve it in the first few distributed trials.

A preliminary and a main experiment were conducted. In the former, *S*s were assigned on a nonrandom intact-group basis. The results of

[1] The authors are greatly indebted to Mr. D. G. Brown, Dr. J. E. Greene, Mr. W. C. Miller, and Mr. G. L. Thornton who helped with the collection of the data. Without this assistance the study would not have been possible. The manuscript has profited from the critical reading by Mr. W. C. Miller.

this study were sufficiently suggestive to justify a more extensive investigation, with randomly assigned groups working simultaneously in different testing rooms.

Preliminary Experiment

Method

Subjects.—The *Ss* were 185 students enrolled in eight psychology classes at the University of Denver. They were assigned to three experimental conditions as intact classes.

Procedure.—All tests were administered on a group basis. The *Ss* of Group I were instructed that they would be given, in turn, two mimeographed reasoning problems to solve. On the first problem, herein referred to as the "Professor Problem," they were allowed 15 min. of massed practice. Correct solutions were accepted even though *S* may have needed more than one attempt to arrive at the correct answer. Guessing was discouraged by telling *Ss* that the number of attempts would be a factor in their score, and that incorrect attempts would be recorded by a tally on their paper. Such tallies were actually made in ink, but were not considered in scoring the problems. As each *S* solved the problem, his paper was collected and the time required for solving was recorded. The purpose of this first problem was to furnish a method for equating the three groups on a task which presumably would be correlated with the experimental problem. Immediately after the "Professor Problem," the second, or "Senator Problem," was passed out and *Ss* were allowed 20 min. of massed practice under identical procedures.

Group II followed the same procedure with the exception that they received distributed practice on the "Senator Problem." They were instructed that they would have 1 min. to work on the problem, and then a 30-sec. rest interval, after which they would have another minute to work on the problem. They were told that this procedure would be repeated several times. The *Ss* were requested to do their best to avoid practicing during rest intervals. Twenty trials were given. Group III followed an identical procedure with the exception that the rest intervals were 4 min.

When several *Ss* raised their hands at approximately the same time to indicate attempted solutions, the times were immediately jotted down by *Es* in order to have an accurate measure for answers proving correct. Although in the distributed conditions *Es* often checked some answers during the first part of rest intervals, there was apparently no greater chance of overlooking *Ss* offering solutions in the massed condition than in the distributed ones. The *Ss* were required to keep their papers face down during rest intervals, even while waiting to be checked. Close observation was maintained to prevent *Ss* from receiving help from neighbors. All indications were that this check was equally effective for each group.

Problems.—Both problems involved deductive processes and were of a type frequently encountered in puzzle anthologies. It is not suggested, however, that the problems were strictly deductive or even that deductive elements could be separated in any clear way from the other thought processes involved. The two problems were specifically constructed for the purposes of this study in order (*a*) to insure closer agreement with the criteria listed earlier, and (*b*) to obviate the possibility of *Ss*' familiarity with materials.

The nature of the problems, and what is here meant by "deductive processes," can best be understood by example. Accordingly, the "Senator Problem" is given below.

> In Washington, D. C., Senators Smith, Johnson, Jones, Miller, Jackson, and Robinson represent, but not respectively, the states of Oregon, California, Nevada, Idaho, Arizona, and Utah. All six of the senators live in the same hotel, but each resides on a different floor. The hotel has six floors.
>
> The senators from Arizona and Utah are good friends and live on adjacent floors. The senator from Utah occupies the higher of the two adjacent floors.
>
> Smith currently has a bill before the Senate and often goes down to the fourth floor to the Idaho senator's room to discuss the bill.
>
> Jones must go up to get to any of the other senators' rooms.
>
> Sometimes Jackson walks up four floors from his room to see the senator from Nevada. When Smith goes down to see Idaho's senator, who lives two floors above the senator from Utah, he never stops by to see Johnson.
>
> Miller and the senator from Idaho belong to opposing political parties, but they both go up sometimes and talk with the senator from Oregon.
>
> Match each senator with the state he represents. (The *S* is provided with a list of senators in the order given first above.)

Results

The numbers of *Ss* solving both problems were: Group I, 22; Group II, 12; and Group III, 5. The total *N*'s

in these groups were 85, 72, and 28, respectively. For Groups I and II the rho's between time scores on the two problems were +.29 and +.13. The scatter plots showed no apparent curvilinearity. Using Old's table of critical values of rho (3), it is apparent that one cannot reject at the .05 level the hypothesis of independence of the variables. In order to obtain estimates of significance of correlation based on all *Ss* in a given condition, fourfold tables on passed-failed information for each problem were constructed for each group. Chi squares of 2.77, 3.29, and 2.05 indicated no significant relationship at the .05 level. In the light of these results it was not feasible to use the "Professor Problem" to match the groups. Although the degree of homogeneity in the *Ss* tested is perhaps related to the lack of significant correlations, it is interesting to note that there was, nevertheless, great intersubject variability on both problems. Recent work in concept formation (9) has also yielded very low correlations for tasks purposely selected to have high communality.

The "Senator Problem" yielded categorical information with respect to whether or not the problem was solved and also offered time scores for those solving it. Percentages of *Ss* solving in Groups I, II, and III were 42, 25, and 42, respectively. Chi square revealed a superiority of Group I over Group II significant at the .05 level, and a difference between Groups II and III, in favor of the latter, significant at the .10 level.

Main Experiment

Method

Subjects.—The *Ss* were 300 students in psychology classes at the University of Denver. Stated ages were obtained from all but four of these. The over-all mean and *SD* were 21.2 and 5.5 yr., respectively, with the individual groups not significantly different at the .20 level. Precautions were taken to avoid testing a second time those *Ss* who were registered in more than one of the classes used in this or the preliminary experiment. The 134 men and 166 women were separately assigned at random to the three experimental conditions. In the case of two *Ss*, where a possible exchange of information was suspected, it was decided to eliminate their papers before knowing whether they would solve. Absences on the experimental day, after groups had been assigned, explain the remaining departure from equal *N*'s in the various groups.

Procedure.—With the following exceptions, the procedure for this experiment was the same as in the preliminary study.

1. Group III had a 90-sec. rest interval between trials. This yielded the most extreme degree of distribution that could be studied in the framework of a 50-min. class period.
2. Due to the unserviceability of the "Professor Problem" as a matching device, only the "Senator Problem" was used.
3. Introspective data concerning practice during rest intervals were obtained after problems had been collected. This information was gathered from the last 108 *Ss* randomly assigned to the two distributed conditions. They were told that it was realized that they might have found it difficult to follow instructions not to practice. It was explained that it would be valuable if they would indicate (*a*) whether they were aware of having practiced during rest intervals, and (*b*) if so, whether the practice had seemed to help.

Results

Table 1 indicates that the 30-sec. distribution group performed significantly less well than either of the other groups, as measured by percentage solving. The difference between the massed and the 90-sec. conditions was not significant. In terms of solution times the three groups did not differ significantly. The Bartlett test for homogeneity was not significant and the *F* ratio was less than unity.

Of the 54 *Ss* questioned in each distributed condition, many reported practice during the rest intervals. Smaller numbers estimated that this practice had helped. The table of Mainland and Murray (8) indicated that Groups II and III differed signifi-

TABLE 1

PERCENTAGES OF *Ss* SOLVING THE PROBLEM UNDER DIFFERENT CONDITIONS—MAIN EXPERIMENT

Degree of Distribution	*N*	Time to Solve		% Solving	x^2	p
		Mean	*SD*			
Massed	102	841.9	170.5	42	6.17	<.02
30-sec. rest	98	873.1	192.2	26	4.12	<.05
90-sec. rest	100	883.1	214.2	39		

cantly in terms of reported rest interval practice (cf. Table 2). Reports that practice had helped did not differ significantly for the two groups at the .05 level.

It was possible to compare the groups in terms of frequency of attempted solutions by counting the ink tallies *Es* had made on the papers. Considering separately those *Ss* who solved, and those who did not solve the problem, the obtained frequencies show in each case a trend toward increased submitting of incorrect answers with greater distribution of practice. The chi square for frequency of submitting one or more incorrect solutions by those who did not solve is significant at the .001 level. For *Ss* who solved, chi square cannot be computed from the threefold table because of small expected frequencies. Therefore, the results were condensed into a 2 × 2 table comparing the massed condition with the two distributed conditions combined. Fisher's exact treatment of a 2 × 2 table yields in this case a p of .006 (cf. Table 3).

TABLE 2

REPORTS OF *Ss* CONCERNING PRACTICE DURING REST INTERVALS

Degree of Distribution	*N*	No. Admitting Practicing	No. Saying Practice Helped
30 sec.	54	20	4
90 sec.	54	32	8
p		<.05	>.05

DISCUSSION

Even though the preliminary experiment used intact classes and the main experiment random groups, percentages of *Ss* solving are consistent in terms of direction of group differences and also in degree. The fact that there were significant differences in terms of frequency of correct solutions, but not in terms of time scores, seems related to the suggestion that varying degrees of distribution are less sensitively reflected by rate of learning than by other measures (2). Other comparisons discussed below will refer to the main experiment unless the sense of the statement dictates otherwise.

Differences in methodology between this and related studies should be mentioned. For one thing, the phase of problem solving at which distribution of practice was introduced, makes the present results quite uncomparable to pre-

TABLE 3

FREQUENCIES OF *Ss* SUBMITTING ONE OR MORE INCORRECT ANSWERS

Condition	*Ss* Not Solving		*Ss* Solving	
	N	No. Giving Incorrect Answers*	*N*	No. Giving Incorrect Answers**
Massed	59	18	43	3
30 sec.	73	30	64	18
90 sec.	61	44		

* $x^2 = 22.9$ ($p < .001$).
** $p = .006$.

vious ones. Furthermore, the decision to study distribution prior to solution necessitated an unusual design for spacing practice, where *S*s in a group varied considerably in the number of practice sessions received. The number of trials for an *S* depended on when he solved the problem. It is interesting to note that even the most proficient *S*s in the two distributed conditions did not solve until their eighth practice trial. The total working times for fastest solvers in the 30-sec. and 90-sec. conditions were 465 sec. and 479 sec., respectively. The most rapid solution for the massed condition was 510 sec.

Another variation between this and most other studies is that the massed condition had no timed interruptions such as are usually obtained in the pauses between successive trials. A worth-while dimension to add to an investigation of this type would be the effect of very brief interruptions of the massed condition (e.g., 20 2- or 3-sec. pauses). It would be desirable to use this and other treatments to shed some light on possible influences of frustration, for example. Behavior of *S*s while working on the problem and their questions, even days later, pointed to extremely high motivation. As might be predicted, practice interruptions in the distributed groups seemed to produce marked frustration. Following experimental sessions many *S*s spontaneously remarked that they had felt these interruptions to be a real handicap. After observing these signs of frustration one might have been tempted to predict an even larger handicap from distribution. Although differences were obtained, perhaps more striking than these are the marked degrees of overlap between conditions, which necessitated quite large *N*'s for the demonstration of differences significant at the levels cited.

The difference between the introspective reports of the 30- and 90-sec. groups is not out of keeping with accounting for the performance differential in terms of practice during rest intervals (Table 2). At the same time, these results do not constitute a verification of such a hypothesis. In planning the study it was decided to consider possible practice during the interval as an integral feature of the dimension of distribution. Basic to this decision was the difficulty in differentiating between simple prevention of practice on the one hand, and inhibitory or facilitating effects on the other, which could have resulted from activity used to fill the interval. The second criterion used in selecting the problem to be solved was for the purpose of at least limiting effective practice during rest intervals. Our early anticipation that *S*s could report very little practice during rest has been shown clearly in error by the above finding.

Another aspect of rest interval activity was an apparent tendency in certain *S*s in the distributed groups to raise their hands more frequently upon being notified of termination of work periods than would be expected from a regular distribution of such responses over the interval in question. However, the greater use of rest periods for checking answers does not necessarily indicate that the distributed groups used less of the work time for checking than did the massed, since the former groups submitted more answers. The possibility of a slight difference between conditions here appears negligible in the over-all interpretation, in view of the fact that mean solving times were numerically, though not significantly, greater with increased distribution. It would have been possible to prevent the use of rest time for checking, by instructing *S*s in the distributed conditions that their answers would not be examined once a rest interval had started. This, however, might have created more serious problems in terms of distraction and frustration.

The trend showing more frequent submitting of incorrect answers with increased distribution may be understood partly in terms of the tendency of some *S*s to offer solutions when notified of the start of rest intervals. However, this does not necessarily completely account for the results. There were 22 more *S*s in Group III than in Group II submitting one or more incorrect answers, whereas both *E*s present in all testing of the 90-

sec. condition recalled that no more than 2 or 3 *S*s had offered answers later than the first 30 sec. of the rest intervals. Among other possible influences, it may be that since total experimental time increased as distribution became greater, this allowed for more forgetting of the proposed penalty for guessing. More complete investigation of the factors involved here is needed.

One of several hypotheses discussed by Underwood (10) to account for Cook's (1) and Ericksen's (4) results is the plausible suggestion that the advantage for massing may have been due to the very long distribution intervals. If this were the case, it would be unnecessary to postulate, as do several writers, a different distribution of practice curve for problem solving from what has usually been found in other learning situations. The present results clearly establish an advantage for massing as compared to a very short distribution interval.

Summary

The effects of distribution of practice on a rational problem were studied. In the main experiment 300 college students each received a total of 20 min. of practice under one of the following conditions: massed; 60 sec. work, 30 sec. rest; and 60 sec. work, 90 sec. rest. Distribution of practice occurred prior to first solution of the problem. The results indicated the following conclusions.

1. In terms of number of *S*s solving the problem, the 30-sec. distribution group was inferior to both the massed and the 90-sec. conditions. The massed and 90-sec. groups were not significantly different.

2. Time scores for those *S*s solving did not differentiate between groups.

3. Two deductive-type problems used in the preliminary study specifically because of their subjective similarity showed no significant relationship in subject performance in any of the groups.

4. There was a significant positive relationship between length of rest interval and frequency of *S*s in a group who offered one or more incorrect problem solutions. This held true both for those who solved and those who did not solve the problem.

References

1. Cook, T. W. Massed and distributed practice in puzzle solving. *Psychol. Rev.*, 1934, **41**, 330–355.
2. Deese, J. *The psychology of learning*. New York: McGraw-Hill, 1952.
3. Dixon, W. J., & Massey, F. J., Jr. *Introduction to statistical analysis*. New York: McGraw-Hill, 1951.
4. Ericksen, S. C. Variability of attack in massed and distributed practice. *J. exp. Psychol.*, 1942, **31**, 339–345.
5. Garrett, H. E. Variability in learning under massed and spaced practice. *J. exp. Psychol.*, 1940, **26**, 547–567.
6. Kendler, H. H., Greenberg, A., & Richman, H. The influence of massed and distributed practice on the development of mental set. *J. exp. Psychol.*, 1952, **43**, 21–25.
7. Leeper, R. Cognitive processes. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 730–757.
8. Mainland, D., & Murray, I. M. Tables for use in fourfold contingency tests. *Science*, 1952, **116**, 591–594.
9. Oseas, L., & Underwood, B. J. Studies of distributed practice: V. Learning and retention of concepts. *J. exp. Psychol.*, 1952, **43**, 143–148.
10. Underwood, B. J. *Experimental psychology*. New York: Appleton-Century-Crofts, 1949.

(Received March 28, 1953)

Journal of Experimental Psychology
Vol. 46, No. 6, 1953

AMOUNT SET AND THE LENGTH-DIFFICULTY FUNCTION FOR A SELF-PACED PERCEPTUAL-MOTOR SKILL[1, 2]

CLYDE E. NOBLE

Perceptual and Motor Skills Research Laboratory, Human Resources Research Center

A number of studies of the phenomenon of *amount set* in human motor performance have indicated an inverse relationship between initial response strength or rate and the anticipated magnitude or length of the activity (**3, 4, 10**). Descriptively, effortful situations seem to elicit energy-conservation behavior on the part of *Ss* confronted with increasing outlays of work.

Previous research has utilized two classes of activity: (*a*) apparatus requiring substantial muscular effort, such as a dynamometer (**10**) and a manual crank (**3**); and (*b*) computational problems, like simple addition (**1, 2, 5**). The "motor" studies have shown consistent results, but when the "mental" skill was investigated over short ranges there was a reversal of the usual effect. Paradoxically, it is this apparent exception which constitutes the classical Bills-Brown experiment (**2**), from which the phenomenon derives its name.

The present experiment is concerned with the effect of extensive variation in quantitative sets on performance on a *self-paced* psychomotor apparatus, the Complex Coordination Test (cf. **7**). It is known (**11**) that skill on this device is relatively unimpaired by work or inhibitory factors as compared, say, with the manual crank apparatus (**3**). Less effortful, *paced* situations, such as the Rotary Pursuit Test and the Two-Hand Coordination Test (cf. **7**), also produce greater response decrements than does the Complex Coordination Test. Hence, the present apparatus should provide a limiting case for what we may term the effort or energy-conservation hypothesis of set effects.

The Complex Coordination Test is well suited to an analysis of perceptual-motor skill into unitary task components. The writer has recently developed a scale of pattern difficulty (**8**) and identified some of the relevant physical variables (**9**) in performance on this device. These studies facilitated the design of the present study.

It was shown that the requisite skill is primarily a matter of imitating each of the three display "movements" with the appropriate control. Given task presentation by paired associates with an invariant sequence, it was reasonable to find movement amplitude as the primary source of difficulty. Furthermore, since the control-display relationships are, except for the elevator, positively correlated or "natural" (previously acquired), it follows that little habit interference should develop among S–R components in Complex Coordination tasks of varying length. Characteristic of such a "noncompetitional" skill, therefore, one would expect a *linear* length-difficulty relationship. This theoretical function, it may be noted, is distinct

[1] The experimental work for this study was performed as part of the United States Air Force Human Resources Research and Development Program. The opinions or conclusions contained in this report are those of the author. They are not to be construed as reflecting the views or endorsement of the Department of the Air Force.

[2] S/Sgt. R. B. Card supervised the experiment and the statistical analyses.

from the *curvilinear* one typically observed in verbal and maze learning situations (cf. 6) in which the successive acts in a serial chain compete with each other for order of appearance. Sequential learning problems, internally considered, are "competitional" in nature.

The purpose of this experiment is twofold: (*a*) to test the effect of a wide range of amount sets on initial performance, and (*b*) to determine a provisional length-difficulty function for psychomotor skills of this type.

Procedure

Apparatus.—Model E of the standard AAF Complex Coordination Test CM 701 (7) was modified to provide detachable connections among the three signal banks. Briefly, the apparatus consists of three parallel banks of 13 red and green signal lamps, a control stick for aileron (A), elevator (E), and a rudder bar (R). The *S*'s problem is to align or "match" successively presented A-E-R patterns of red (stimulus) lamps with green (response) lamps by means of appropriate stick and rudder manipulations. When a match is correctly made and maintained for .5 sec., a stepping relay advances automatically to a new A-E-R pattern. This procedure is repeatable indefinitely for any desired length of task.

Forty stimulus patterns of known absolute difficulty (8) were arranged in a random sequence. An accessory selector switch permitted any number from 5 to 40 consecutive patterns to be presented and terminated automatically. Alternately-energized Standard Electric clocks (S-1) recorded the time per match for individual A-E-R combinations, and rest periods of 2 min. were given between tasks. Calibration procedures regarding time delay, effective scoring area, control travel, and spring compression were followed as described previously (9).

Subjects.—One-hundred sixty basic airmen, split equally and without bias into experimental and control groups, comprised the sample. None of the *S*s reported prior experience with the task.

Method.—Eight task lengths were investigated: 5, 10, 15, 20, 25, 30, 35, and 40 matches. The experimental design utilized a basic 8 × 8 treatments × *S*s cycle ($N = 8$), in which the interaction of treatment (task length) and order (facilitative or decremental) effects was confounded with individual differences. A-E-R pattern difficulty was also equalized among task conditions, thus permitting unbiased comparisons of the treatment means.

This basic counterbalanced design was then replicated ten times for each group ($N = 80$). The two groups differed only in that the experimental (set) *S*s were informed of the exact number of matches (35, 10, etc.) that awaited them, whereas the controls (no-set) were not given this information. Modified AAF instructions (cf. 7) were provided, followed by two demonstration matches by *E*. Each *S* was then permitted to make five matches for familiarization purposes.

Results and Discussion

Figure 1 shows a comparison of first responses in the experimental and control groups for each of the eight tasks. The response measure is the mean of the median initial match scores for the ten replications. Complex Coordination scores are characterized by skewed, multimodal frequency distributions. However, means of subsample medians have been found empirically to be linearly related to general medians, which have a fairly symmetrical distribution. First responses are taken since the principal effect of amount set has been observed to be an initial one.

An 8 × 2 analysis of variance was set up for the data in Fig. 1, in which the two row effects are confounded with group sampling differences. For the purpose of trend tests, each cell is considered to contain ten measures,

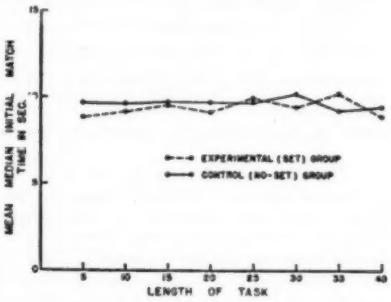

Fig. 1. The relationship between amount set and length of task for initial performance

where each score is actually a composite (median) of eight raw scores with order effects completely counterbalanced. A summary is presented in Table 1.

The null hypothesis that the trends of Fig. 1 are parallel is tested by $\sigma^2_{R_E C}/\sigma^2_{error(w)}$, giving an F of 1.05 which is not significant ($p > .05$). In other words, the differences between the control and experimental groups are equal except by chance for all values of anticipated task length. Therefore, there is no interaction between task length and amount set.

The hypothesis that the two groups in Fig. 1 coincide is evaluated by $\sigma^2_R/\sigma^2_{error\ (b)}$, which yields an F of 1.29. This ratio is also not significant ($p > .05$), indicating that the effect of set averaged over all conditions of task length is negligible.

Finally, a test of "no trend" in the average effects of task length is given by $\sigma^2_C/\sigma^2_{error\ (w)}$. This ratio yields an F less than 1.00, which is consistent with the hypothesis that initial performance is equal for all tasks regardless of amount set. By sustaining the null hypothesis in this case we have incidentally double-checked the assumptions regarding the equalization of order and pattern effects.

Summarizing the foregoing analyses, there are no significant trends in the initial match data attributable either to amount set, task length, or to their interaction. The data pertaining to the second, third, and fifth responses were also examined by the same technique. Although the variability was greater, due to the sequential nature of the matches (8), the general results were the same.

TABLE 1

ANALYSIS OF VARIANCE OF INITIAL RESPONSE MEASURES

Source of Variance	*df*	Mean Square	*F*
Between *Ss*			
Set	1	3.5730	1.29
Error (b)	18	2.7727	
Within *Ss*			
Task length	7	1.3033	<1.00
Set × task length	7	1.8423	1.05
Error (w)	126	1.7476	
Total	159		

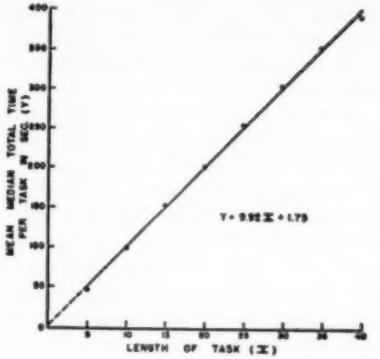

FIG. 2. The relationship between task difficulty and length of task

Figure 2 presents the length-difficulty function. The relationship is linear, as shown by the descriptive least-squares equation,

$$Y = 9.92X + 1.75. \qquad (1)$$

Extrapolation to the ordinate indicates a rational zero origin, since the average residual was 3.30 sec., which exceeds the intercept value of 1.75. Practically as good a fit can be obtained by starting the line at the origin and using a slope based on the average of all single matches. In this case the equation becomes

$$Y = 9.955X. \qquad (2)$$

On the ordinate of Fig. 2 is plotted mean median total task time, while the abscissa indicates length of task as in Fig. 1. The data for the experimental (set) and control (no-set)

groups were pooled to give a single, averaged function. Thus, each point from task length 5 to 40 represents the performance of 160 *Ss*.

A trend analysis need not be carried out for the length-difficulty data since the slope is obviously significant. It is instructive, however, to test the goodness of the fit in Fig. 2. This is provided by the ratio $\sigma^2_d/\sigma^2_{error\ (w)}$, where σ^2_d is the variance estimate for the departure of the experimental points from the fitted linear regression line, and where the error term has the same meaning as that in Table 1. Since we have computed two empirical constants in Equation 1, σ^2_d will be based on 8 *df* less 2. For 6 and 126 *df*, the resulting *F* is less than unity, indicating an excellent fit. Additional information is provided by the coefficient of determination r^2, computed between the observed and calculated *Y* values. The proportion of total variance in task difficulty accounted for by the linear equation is .999.

The findings shown in Fig. 1 and Fig. 2, therefore, verify the two hypotheses advanced at the outset: (*a*) the effect of amount set on initial performance is negligible; (*b*) the length-difficulty function is linear. From the present viewpoint, these results are consistent with previous research. In general, it is proposed that psychomotor skills involving considerable physical work will show an interaction between conditions of quantitative set and variations in length of task. Furthermore, as required effort is increased the interaction should rise in significance.

This effect should also be revealed in the relationship between difficulty, defined by total task time, and task length, defined by number of components. The length-difficulty function is expected to originate at zero and to be linear over a wide range, but the slope parameters should vary positively with the amount of work involved. Variations in required effort are theoretically manipulable by re-engineering the physical work (force × distance) characteristics within the components of the task.

An example of such alteration in the case of the Complex Coordination Test would consist in increasing the spring compression or the arc travel of the apparatus controls. With A, E, and R movements more heavily loaded or greater in amplitude, the slope of the length-difficulty function should exceed the standard task value of 9.92 determined in Equation 1. Judging from previous research (9), moreover, changes in amplitude requirements will prove more important than any modification of the force loadings.

As a final note, the *within-components* source of variance should be distinguished from the *among-components* type. The former is thought to derive primarily from the *work* characteristics of a motor skill, while the latter mainly reflects its *complexity* mediated by habit interference. Work variations are expected to affect the steepness but not the linearity of the length-difficulty function. Variations in complexity should change both slope and form. Under conditions of serial learning, for example, positive acceleration in difficulty with lengthening tasks is commonly reported (cf. 6). The present position regards curvilinearity as a special characteristic of a class of activities which may be called "competitional" in nature. In comparison, the theory of "noncompetitional" skills should be relatively simple.

Summary

The relationship between amount set and length of task in self-paced psychomotor performance was investigated in a low-work, "noncompetitional" skill which would, by hypothesis, (*a*) be immune to variation in quantitative set, and (*b*) increase in difficulty proportionally with length of task.

Eight task lengths were studied: 5, 10, 15, 20, 25, 30, 35, and 40 matches, with 2-min. rests between successive tasks. One group ($N = 80$) was informed of the exact number of matches to be made during each task period, another ($N = 80$) was not given this information. The results, in conformity with hypothesis, were:

1. Analysis of variance tests of trend revealed no significant patterns in the *initial* performance

data attributable either to amount set, task length, or to their interaction.

2. When difficulty (Y), defined by *total* task time, was plotted against length of task (X), defined by number of matches, a linear function was obtained.

It was hypothesized that perceptual-motor skills involving greater physical work requirements than the one investigated should not only reveal amount set–task length interaction, but should also be characterized by length-difficulty functions of increased slope. Unless there is substantial habit interference among the components in a given situation, difficulty is expected to be proportional to length of task, with the empirical slope parameter serving as an index of its effortfulness.

References

1. Barmack, J. E. The length of the work period and the work curve. *J. exp. Psychol.*, 1939, **25**, 109–115.
2. Bills, A. G., & Brown, C. The quantitative set. *J. exp. Psychol.*, 1929, **12**, 301–323.
3. Bilodeau, I. McD. Performance of an effortful task with variation in duration of prior practice and anticipated duration of present practice. *J. exp. Psychol.*, 1953, **46**, 146–153.
4. Katz, D. Gestalt laws of mental work. *Brit. J. Psychol.*, 1949, **39**, 175–183.
5. Krueger, W. C. F. The influence of amount limits and time limits upon the rate of work. *J. appl. Psychol.*, 1937, **21**, 113–118.
6. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.
7. Melton, A. W. (Ed.) *Apparatus tests.* AAF Aviation Psychology Program, Research Report No. 4. Washington: U. S. Govt. Print. Off., 1947.
8. Noble, C. E. A representative scale of difficulty in the Complex Coordination Test (CM 701 E). Hum. Resour. Res. Cent., *Res. Bull.*, 1952, No. 52–41.
9. Noble, C. E. Some physical sources of difficulty in the Complex Coordination Test (CM 701 E). Hum. Resour. Res. Cent., *Res. Bull.* 1952, No. 53–7.
10. Ross, S., & Bricker, P. D. The effect of an amount-set on a repetitive motor task. *J. exp. Psychol.*, 1951, **42**, 39–43.
11. Spence, K. W., Buxton, C. E., & Melton, A. W. The effect of massing and distribution of practice on the SAM Complex Coordination Test. Washington: Civil Aeronautics Administration, Division of Research, Report No. 53, 1945.

(Received April 7, 1953)

Journal of Experimental Psychology
Vol. 46, No. 6, 1953

TRANSFER OF TRAINING IN MOTOR LEARNING AS A FUNCTION OF DISTRIBUTION OF PRACTICE[1]

VICTOR E. MONTGOMERY[2]

Northwestern University

The purpose of the present study was to investigate transfer from a first to a second motor task as a function of distribution of practice on both tasks. The study stems from a previous one (2), employing the same tasks, in which it was shown that positive transfer increased directly both with degree of learning of Task I and with degree of formal similarity between tasks. Although there has been much research on distribution of practice as a variable in learning (6), there is no information available as to the effect of this variable on transfer.

EXPERIMENT I

Method

Apparatus.—The *Ss* were required to learn six perceptual-motor paired associates. The stimuli were six different hues presented in random order on the same ground-glass screen. The response was the movement of a lever, held by the right hand, into any one of six radial slots. Various S–R combinations could be arranged by means of telephone jacks and sockets. Simultaneously *S* held another lever with the left hand, which was essentially a steadiness test. Correct performance of the right-hand task was contingent upon correct performance of the steadiness test. The apparatus was subject-paced, i.e., *S* could work as fast as he liked. The apparatus was set to provide, automatically, 20-sec. trials and 10-sec. intertrial rests. Separate recording of four response measures was possible: correct responses, shallow errors, and deep errors with the right-hand lever, and steadiness time with the left lever. A complete description of the apparatus may be found in the earlier study (2).

Experimental design.—The experiment was a 3×3 factorial design with three degrees of distribution of trials over days on both Task I and Task II. All *Ss* were given 36 trials on each task. The interval between tasks was 24 hr. The three degrees of distribution of trials over days for both tasks, and the group notations for Task I are as follows: 36 trials given all on one day (D–I); 18 trials a day on two successive days (D–II); 12 trials on each of three successive days (D–III). The notations for the comparable Task II distributions are as follows: D–A, D–B, and D–C. Fifteen *Ss* were assigned to each of the nine cells.

Tasks.—For Task I the lights were paired with the slots in 15 different combinations. Each *S* in a cell of 15 *Ss* was assigned one of these combinations. The same 15 combinations were used for each of the nine cells of *Ss*. For each of the 15 combinations another combination for use on Task II was made up by newly pairing each light with a different slot. Again, each *S* in a cell was assigned one of these Task-II combinations and the same 15 combinations were used for all cells on Task II. The 15 combinations for both tasks were chosen from those used in the earlier study (2). As was described in that study, both tasks can be considered equal in difficulty, because of the method of choosing the combinations. Thus no control groups were run; data from Task I served as control for Task II.

Subjects and procedure.—The *Ss* were 135 males enrolled in the introductory psychology course at Northwestern University. They were rewarded with one point of examination credit for each day served in the experiment. In addition, they were told that the experiment was being done for the Air Force. From *Ss'* comments it appeared that motivation was high.

When *S* first appeared for the experiment he was given a crude color-discrimination test which

[1] This is the second in a series of articles on transfer in motor learning. The research was performed under Air Force Contract No. 33 (038)-11396 between the Aero Medical Laboratory of the Wright Air Development Center, Air Research and Development Command, and Northwestern University. The contract was under the direction of Dr. C. P. Duncan and Dr. B. J. Underwood. The writer wishes to express his appreciation to the contract directors for their valuable suggestions and advice during the course of the study.

[2] Now at 6561st Research and Development Group (Perceptual and Motor Skills Research Laboratory), Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

consisted of calling out the names of the six stimulus lights as they were presented in random order. Following color naming, *S* was asked to take his place in front of the apparatus with his hands on the levers. Then the height of the table carrying the apparatus was adjusted until the tops of the levers were approximately 1 in. below elbow height. Since *S* stood up for all practice trials, this height seemed to be least fatiguing.

The purpose of the experiment was explained in very simple terms, using everyday situations, e.g., driving two different makes of automobiles. A rather lengthy set of instructions had been previously worked out. The gist of the instructions was as follows: *S* was told that he was to learn, by moving the right-hand lever into the slots, which light was connected to which slot, but it was emphasized that this was not sufficient; he was to turn off as many lights as possible during the 20-sec. trials. He was informed that during the 10-sec. rest periods he would be told how many correct responses he had made, i.e., how many lights he had turned off during the previous trial.

On the day *S* was to begin Task II he was told that now the lights and slots were connected in a different way and that he was to go ahead and learn this task in the same way he had the first one.

Results

Although four response measures were recorded, only two of these, correct responses and shallow errors, will be used. It was shown elsewhere (2) that little additional information is contributed by the deep errors or the time scores.

Task I.—Figure 1 shows performance (correct responses) for the three groups (45 *Ss* each) with different distributions of practice on Task I. Breaks in the curves for Groups D–II and D–III indicate 24-hr. rests.

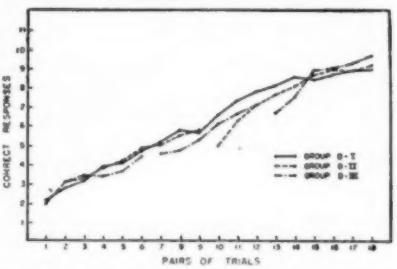

FIG. 1. Performance on Task I for groups differing in distribution of trials over days

Although the curves in Fig. 1 indicate little or no difference in performance as a function of practice distribution, tests of significance were performed on the means for the following blocks of trials: 1–12, 13–24, 25–36, 1–18, 19–36, and 1–36. Because of heterogeneity of variance, the *t* ratio by an approximation method (3) was used.

Since the groups were treated identically during Trials 1–12, comparing the groups for this block will indicate whether the original assignment of *Ss* resulted in comparable groups. Since none of the *t* values for this or any other block of trials was significant, it is clear that the three groups did not differ significantly either from assignment of *Ss* to groups or as a result of distributing practice trials over days.

The error data were analyzed in the same way with the same results; no significant differences were found.

Task II.—Since distribution of practice was found to have no effect on Task I, we shall, for a first analysis, combine the subgroups D–I, D–II, and D–III within each of the Groups D–A, D–B, and D–C. The performance of these groups (45 *Ss* each) on Task II is shown in Fig. 2. Comparison of Fig. 2 with Fig. 1 shows that all groups transferred positively to Task II. However, the curves in Fig. 2 do not show much evidence either of differential performance as a function of distributing practice on Task II, or of differential transfer from Task I. The statistical analysis was the same as before; because of heterogeneity of variance, *t* ratios were computed between the means for each of the six blocks of trials listed

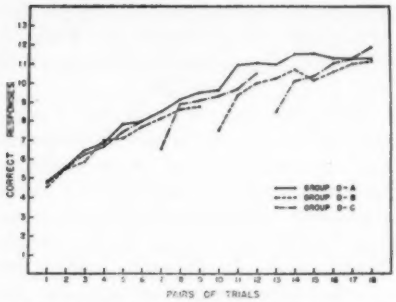

FIG. 2. Performance on Task II for groups differing in distribution of trials over days. Distribution of practice on Task I is held constant.

above. None of the t values for correct responses was significant.

The same analysis on the error data yielded three t values (out of 18) significant ($p = .05$). The groups differed in errors during the last block of 12 trials (Trials 25–36). Throughout both Task I and Task II there were losses in correct responses and increases in errors over 24-hr. rests. These decrements were usually brief; recovery was rapid, except in errors during this last block of 12 trials on Task II. This apparently accounts for the significant differences.

It is possible that the preceding analysis, which involves combining subgroups, does not go far enough toward testing for differential transfer. A further analysis was made by comparing all nine subgroups by nonparametric analysis of variance, a test which does not depend upon homogeneity of variance (5). A separate analysis was made for each of the six blocks of trials for both correct responses and errors. The only significant difference ($p = .05$) was on Trials 25–36 for errors. Thus there is not enough evidence to modify the conclusions previously stated; distribution of practice over days did not result in differential performance on either task or in differential transfer.

Since the results of this experiment were negative, the procedure was reviewed, and the possibility occurred that the 10-sec. rest interval between trials might have been counteracting the day distribution by allowing time for dissipation of possible inhibitory factors built up during the 20-sec. trial. Therefore, additional *Ss* were run with no intertrial rests, i.e., practice was continuous within days.

EXPERIMENT II

Method

With appropriate modifications of the apparatus 30 male naive *Ss* were given 12 min. (36 trials) of *continuous* practice on Task I on one day. This group will be called Group C–I, since it is comparable with Group D–I of Exp. I. Group C–I was divided into two groups of 15 *Ss* each, equated for performance on Task I, for the differential conditions on Task II. One of these groups was given 36 trials of continuous practice on Task II on one day (Group C–A); the other 15 *Ss* had their 36 trials on Task II spread over three days, with 4 min. of continuous practice per day (Group C–C).

The instructions were the same as those in Exp. I, except that *Ss* were told to work continuously until the signal to stop. The number of correct responses made by *S* was again reported to him every 20 sec. A timer independent of the apparatus was used to indicate the 20-sec. intervals.

Results

Figure 3 shows performance on Task I for Groups D–I and C–I, the groups

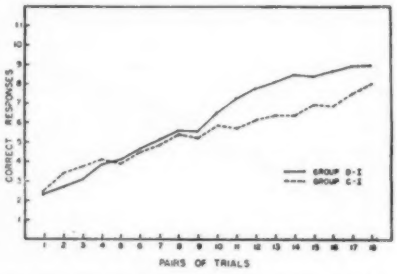

FIG. 3. Performance on Task I for comparable groups from Exp. I and II

from the two experiments that are comparable. The groups appear to differ late in practice, and this is borne out by the significance test; the groups differ at the 1% level over Trials 25–36. Thus the usual result is found that distributing practice by giving intertrial rests yields higher performance. The groups did not differ in errors.

Figure 4 shows performance on Task II for the four groups from the two experiments that can be compared. It will be recalled that Groups D–A and C–A had all of the practice on Task II on one day, while in Groups D–C and C–C the practice was spread over three days. Further, in Groups C–A and C–C there were no intertrial rests.

By comparing Fig. 1 and 3 with Fig. 4 it may be seen that all groups showed positive transfer to Task II. However, there is again little indication of differential transfer in the curves of Fig. 4. Group C–C began Task II at a lower level of performance, but the difference between it and the other three groups at the initial point is not significant.

Again, because of heterogeneity of variance, and unequal N's, the approximation t ratio was used to compare the groups in Fig. 4. Comparisons

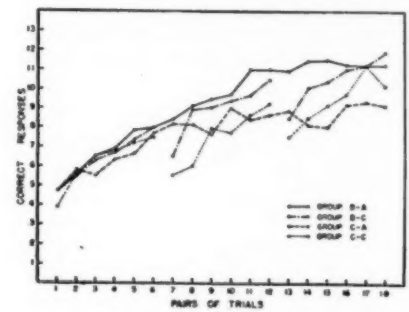

FIG. 4. Performance on Task II for comparable groups from Exp. I and II

TABLE 1

VALUES OF t FOR DIFFERENCES BETWEEN COMPARABLE GROUPS OF EXP. I AND II

Groups Compared	Trials			
	1–12	13–24	25–36	1–36
C–C vs. C–A	.13	.71	.68	.00
C–C vs. D–A	.60	2.93**	2.49*	2.23*
C–C vs. D–C	.46	1.66	1.31	1.32
C–A vs. D–A	.95	1.77	2.57*	2.23
C–A vs. D–C	.31	.59	2.98**	1.32
D–A vs. D–C	.11	1.96	1.68	1.31

* Significant at $p = .05$.
** Significant at $p = .01$.

were made at each block of 12 trials, and over all trials. The t values are shown in Table 1. It may be seen that none of the t values for Trials 1–12 is significant. Thus there is no evidence of differential transfer as a function of time between trials on Task I, at least in this portion of practice on Task II. As Fig. 4 shows, and as is borne out by the significant t values in Table 1, the usual inferior performance of groups working under continuous practice eventually appears. These differences are probably due to differential distribution of practice on Task II itself, rather than evidence of transfer from different conditions on Task I.

The error curves for the four groups show much the same phenomena as the correct-response curves, and will not be presented.

DISCUSSION

The fact that there was no differential transfer as a function of either method of distributing practice in this study raises one problem when the results are compared with those obtained in the previous study (2). In that earlier experiment significant differential transfer to Task II occurred by varying the degree of learning on Task I. In the present study

there is one case (Fig. 3) where groups terminated practice on Task I at significantly different levels of performance. In this case the differential performance was the result of massing practice in one group, whereas in the previous study it was the result of varying the number of trials. Yet these groups shown in Fig. 3 did not show differential transfer (Fig. 4).

Either one of two points may account for the apparent contradiction in the results of the two studies. In the first place, the differences in performance among the groups at the end of Task I were, in general, greater in the earlier study than the difference between the groups shown in Fig. 3. Perhaps the difference in the present study was not large enough to be reflected in transfer.

In the second place, it is possible that in the present study the group working under continuous practice (Group C–I) on Task I showed merely a temporary depression in performance. If there had developed an inhibitory process, such as reactive inhibition (**1, 4**), that soon dissipated, then the groups may have learned the same amount although they performed differently. Unless the time between tasks was very short, so inhibition did not dissipate, one would not expect transfer to reflect temporary depressions in performance. Since a 24-hr. interval separated the tasks, there was plenty of time for dissipation of temporary inhibition.

Summary

Transfer in motor learning was studied as a function of distribution of practice on both Task I and Task II. In Exp. I 36 20-sec. trials with 10-sec. intertrial rests were, in different groups, distributed over 1, 2, or 3 days on both tasks. In Exp. II 36 20-sec. trials with no intertrial rests were given on Task I in one day, and distributed over 1 or 3 days on Task II. The apparatus used was the Duncan-Underwood Star Discrimeter, which required performance on a steadiness test while simultaneously learning six perceptual-motor paired associates. The findings were:

1. In Exp. I there were no differences in performance in terms of correct responses among the groups on either task. There were a few differences among the groups in errors near the end of Task II. Transfer was positive but did not vary as a function of practice distribution.

2. In Exp. II continuous practice resulted, late in practice, in poorer performance on both tasks than that achieved by groups practicing with intertrial rests. There was again, however, no evidence of differential transfer; all groups performed equally well in the early portion of Task II.

It was suggested that continuous practice resulted in a temporary depression in performance, but not in learning.

References

1. Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
2. Duncan, C. P. Transfer of training in motor learning as a function of degree of first-task learning and inter-task similarity. *J. exp. Psychol.*, 1953, **45**, 1–11.
3. Edwards, A. *Experimental design in psychological research.* New York: Rinehart, 1950.
4. Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
5. Moses, L. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122–143.
6. Underwood, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.

(Received April 17, 1953)

Journal of Experimental Psychology
Vol. 46, No. 6, 1953

RETENTION OF TRANSFER IN MOTOR LEARNING AFTER TWENTY-FOUR HOURS AND AFTER FOURTEEN MONTHS[1]

CARL P. DUNCAN AND BENTON J. UNDERWOOD

Northwestern University

In the first (3) of this series of studies on transfer in motor learning the effect on transfer of two variables was investigated: degree of learning of the first or training task (Task I), and similarity between Task I and the second or transfer task (Task II). Although the data were not reported, the experiment also included a test for retention of Task II measured 24 hr. after acquisition. Retention of Task II was measured again during the following year. The present study therefore reports data on retention of Task II measured both 24 hr. and 14 months (average) after acquisition.

Measuring retention of the second of two successively learned tasks is the standard design for studying proactive inhibition of recall (6, 7). Proactive inhibition is usually expected to occur only when the two tasks are mutually interfering, i.e., when there is some evidence of negative transfer from Task I to Task II (7). It has already been shown (3) that the tasks used in the present research transfer positively. And as will be seen, when retention of Task II was measured 24 hr. after acquisition, there was no evidence of proactive inhibition. However, it is assumed that proactive inhibition increases with time (6); thus the results to be reported for retention after more than a year without practice should show whether proactive inhibition occurs even between tasks which transfer positively.

As far as is known, there are no other studies of retention of the second of two successively learned tasks after long intervals without practice. There have been some studies (e.g., 1, 2) in which the interval between original learning of the training and transfer tasks was varied, thereby studying retention of transfer from Task I measured during acquisition of Task II. The present study differs in that retention of transfer is investigated by measuring recall and relearning of Task II.

METHOD

Since this study reports data from a portion of the same *Ss* of the same task described in the earlier paper, details of procedure concerning *Ss*, variables, tasks, and apparatus can be obtained from that report (3). Only the major points of procedure will be given here.

Apparatus.—The *S* learned to move a lever, held by the right hand, into six radially-arranged slots in response to six colored-light stimuli. Simultaneously, *S* held another lever steady with the left hand. If the left lever was not held steadily, correct responses with right lever could not be recorded. Thus, the apparatus required both paired-associates learning and hand steadiness. Each trial was 20 sec., with 10-sec. intertrial rests. Separate counters recorded movements of the right lever as correct responses or errors, and a clock recorded steadiness time for left lever. With this apparatus Task I is a particular pairing of lights and slots which is different for each *S* within a group of 25 *Ss*. Task II is provided by differently pairing the lights and slots. Choice of the light-slot combinations is such that, as explained more fully in the first report (3), the tasks are considered to

[1] This is the third in a series of papers (3, 5) on transfer in motor learning. The research was performed under Air Force Contract No. 33-(038)-11396 between the Aero Medical Laboratory of the Wright Air Development Center, Air Research and Development Command, and Northwestern University. Thanks are due to Dr. V. E. Montgomery, chief assistant on the contract.

be equal in difficulty. This has the advantage that data from Task I can be used as control for Task II.

Conditions.—In the earlier paper data were reported for transfer to Task II as a function of two variables: degree of learning of Task I, and similarity between tasks. There were four degrees of the learning variable: 10, 40, 80, and 180 trials, denoted as Groups I, II, III, and IV respectively. Similarity between tasks was defined by the *number* of light-slot connections that were *different* on Task II from what they had been on Task I. The three degrees of similarity were obtained by newly pairing 2, 4, or all 6 lights and slots for Task II; these are denoted as Groups A, B, and C respectively. Thus in Group A (high similarity) only two lights and slots were re-paired on Task II; the other four remained the same. The variables were studied in a 4×3 factorial design with 300 *Ss*, 25 in each cell.

Two minutes after completion of Task I each *S* was given 60 acquisition trials on Task II. Twenty-four hours later each *S* was given 20 relearning trials on Task II. The 24-hr. retention data for all 300 *Ss* have not been published.

Subjects.—Of the 300 *Ss* originally run, 186 returned for the second test of retention of Task II the following year. When *S* came in for this long-term retention test, he was told that he was going to have exactly the same task, the same light-slot relationship, that he had been working on when he finished the experiment the previous year. Twenty relearning trials were given. Table 1 shows the number of *Ss* in the major groups; each of the 12 subgroups contains 15, 16, or 17 *Ss*, all that were available out of the original 25, for the long-term test. Table 1 also shows, for the major groups, the mean and the range of days elapsed between the 24-hr. and the long-term retention tests. For the total group of 186 *Ss* the mean days was 419.9 (14 months) and the total range of days was 341–562.

TABLE 1

NUMBER OF *Ss* IN EACH MAJOR GROUP, AND THE MEAN AND RANGE OF DAYS ELAPSED BETWEEN THE 24-HR. AND THE LONG-TERM RETENTION TESTS

Groups	*N*	Mean Days	Range of Days
Learning			
I	48	423.9	341–537
II	47	413.7	348–530
III	46	427.5	343–562
IV	45	414.6	348–543
Similarity			
A	62	420.9	350–562
B	63	427.6	348–543
C	61	411.3	341–537

RESULTS

Acquisition of Task II.—The major results are presented in Fig. 1 and 2, where correct responses per trial are plotted for pairs of trials. In Fig. 1 the four Learning Groups are shown with similarity held constant; Fig. 2 shows the three Similarity Groups with learning held constant. From left to right these graphs show:

1. Performance at end of practice on Task I (mean of last two trials), indicated by heavy points. Since similarity was not a variable on Task I, Fig. 2 shows a single point, the grand mean for all *Ss*.

2. Acquisition of Task II. In both Fig. 1 and 2 the lowest curve is the control. As explained earlier, this control curve is plotted from some of the Task I data; specifically, the curve shows performance on the first 60 trials of Task I by Groups III and IV combined ($N = 91$). These were the only groups given 60 trials on Task I, which is the number of trials given on Task II.

3. At the top of the figures, relearning of Task II after 24 hr., and at the bottom, relearning again after an average of 14 months. The lowest curve among those for 24-hr. retention is a control; it was plotted from the scores of Group IV ($N = 45$) on Trials 61–80 of Task I. This group had their first 60 trials on Task I on one day, followed by a 24-hr. rest before continuing with Trial 61. Therefore, their Trials 61–80 on Task I can be used as a control to show, in effect, what relearning of Task II would have been after a 24-hr. rest by a group that had had no Task I. Unfortunately, there is no control for 14-month retention.

The 186 *Ss* on whom the present report is based are closely representa-

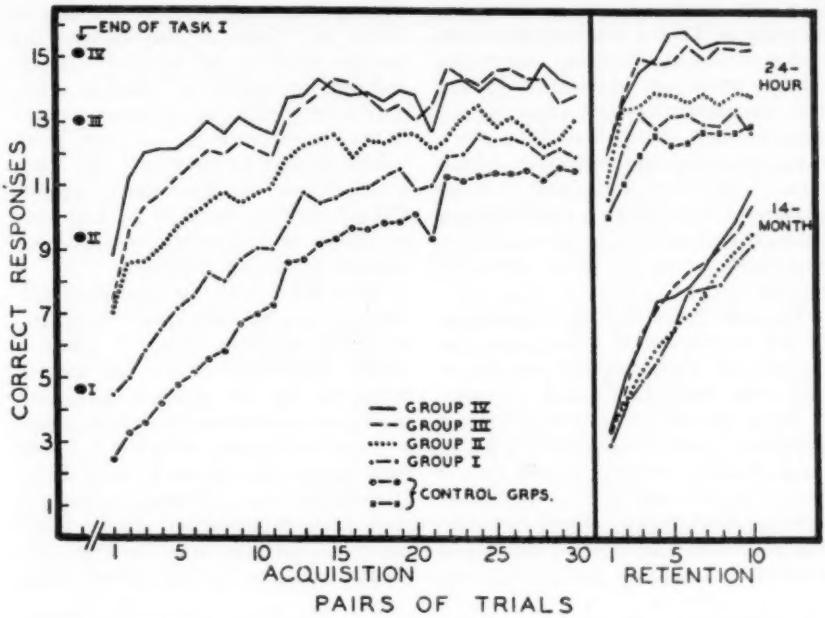

FIG. 1. Acquisition and retention of Task II for the four groups having different degrees of learning on Task I. Intertask similarity is held constant.

tive of the original 300 *Ss*. On performance at the end of Task I and on acquisition of Task II, the curves of Fig. 1 and 2 are highly similar in both slope and absolute level to the results for all 300 *Ss*, as comparison of the corresponding figures (Fig. 2 and 3) in the earlier paper (3) will show. The same is true for 24-hr. retention, although these data for all 300 *Ss* have not been published to permit comparison.

In the earlier paper analysis of variance of the acquisition data showed that either over the first 30 or over all 60 trials of acquisition, both the learning and similarity variables yielded differential positive transfer; the *F* values were highly significant for both variables, but there was no interaction. Since the same results were found upon analysis of the data for the present 186 *Ss*, the analysis will not be reported. Over the last 30 trials of acquisition similarity was not a significant variable for the 300 *Ss* but just reached significance at the 5% level ($F = 3.07$) for the 186 *Ss*.

No error or time scores (steadiness time with the left lever) for acquisition of Task II will be reported, since upon analysis it was found they added nothing to the information derived from analysis of correct responses.

Retention of Task II after 24 hr.— Relearning of Task II is depicted by the curves in the upper right-hand portion of Fig. 1 and 2. It can be seen that there was forgetting in all groups. A measure of recall was obtained by subtracting score on the first relearning trial from mean score on the last ten acquisition trials. These losses are shown in Table 2.

From Table 2 it can be seen that except for Group III, which is out of

line, the amount forgotten increased directly as Task I learning increased and as similarity decreased, suggesting relative proactive inhibition. It may also be seen that all experimental groups forgot more than the control, suggesting absolute proactive inhibition. The recall data will first be analyzed for significance of differences among experimental groups and then comparisons will be made with the control.

The test for differential forgetting among experimental groups was by analysis of variance, the results of which are shown in Table 3. It may be seen that the interaction term is significant, probably due to Group III. No particular importance will be attached to the significant interaction, since the same analysis based upon the original 300 *Ss* did not yield significant interaction and there was no reversal among forgetting scores as appears in Table 2. Table 3 also shows that neither learning nor similarity is significant, regardless of whether interaction or within-groups mean square is used as error (*F* for learning, using within groups as error, is 1.38). We conclude that the experimental groups did not differ in recall of Task II over the 24-hr. rest; there is no evidence for relative proactive inhibition.

Since there were no significant differences among experimental groups, we could combine them to make a single comparison with the control group to test for absolute proactive inhibition. However, the means would be based on greatly different *N*'s, 186 experimental *Ss* versus 45 control *Ss*. It seems better to compare the control with each experimental group separately. This was done, and of the seven *t* tests the only significant value

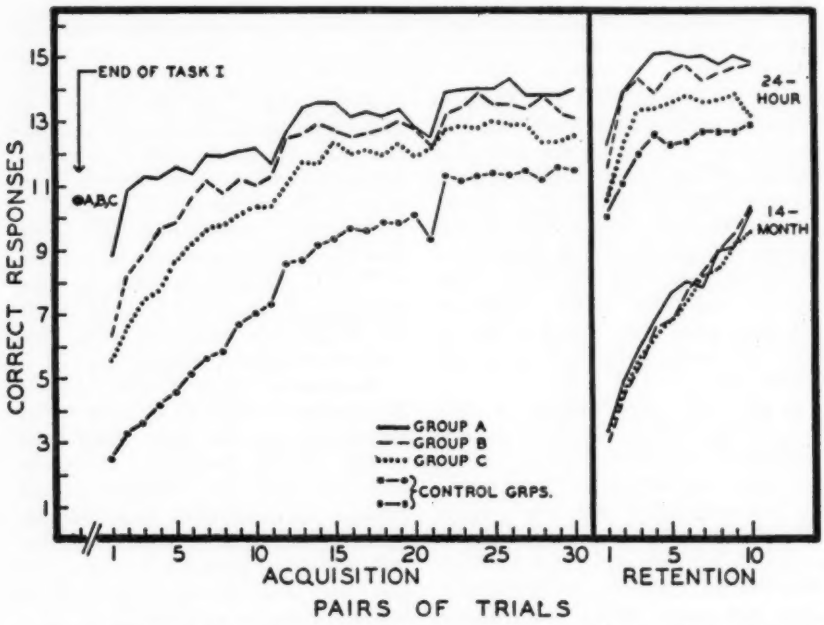

FIG. 2. Acquisition and retention of Task II for the three groups having different degrees of intertask similarity. Degree of Task I learning is held constant.

TABLE 2

RECALL OF TASK II AFTER 24 HR.

Groups	Score		
	Before Rest	After Rest	Difference
Learning			
I	12.09	9.92	2.17
II	12.79	10.32	2.47
III	14.18	10.90	3.28
IV	14.44	11.66	2.78
Similarity			
A	13.88	11.53	2.35
B	13.53	10.71	2.82
C	12.64	9.80	2.84
Control	11.14	9.31	1.83

was that resulting from comparing Group III and the control, where t was 2.14. Again, there seems no reason to consider the greater forgetting in Group III as anything more than a random fluctuation. At the same time it deserves passing mention that it is Group III which reached a level of mastery on Task II where response competition between the tasks might be expected to occur, i.e., mastery equal to or slightly greater than that achieved on Task I. However, there does not seem to be sufficient evidence for absolute proactive inhibition after 24 hr.; we conclude that it did not occur.

Returning to Fig. 1 and 2, it can be seen that all experimental groups continue to be superior to the control, and to differ among themselves during relearning. Analysis of variance among the experimental groups on total score for the 20 relearning trials showed both learning and similarity highly significant, with F's of 7.49 and 5.33, respectively. Interaction was not significant. Thus, there is proactive facilitation, rather than proactive inhibition, from practice on Task I for all degrees of learning and similarity.

Again, analysis of errors and time scores showed little. There was a slight increase in errors in Groups B and C, but in no other groups, over the 24-hr. rest. The time score showed forgetting, but the forgetting reported above in terms of correct responses did not appear to be merely loss of skill with the left lever. The correlation between total correct responses and total time over the 20 relearning trials for all 186 *Ss* was .652; thus about 42% of the variance in correct responses might be due to variance in time score.

Retention of Task II after 14 months. —This is shown by the sets of curves in the lower right-hand corner of Fig. 1 and 2. There is no control curve, but by comparing the 14-month curves with the control for acquisition of Task II, shown in the lower left of the figures, it can be seen that performance on the 14-month test was superior to that of *Ss* being introduced to the apparatus the first time (because this control curve is Task I data). It is not known, however, what performance on Task II after 14 months would have been without practice on Task I.

Recall after the long interval was measured by subtracting score on the first trial of the 14-month test from mean of the last five trials of the 24-hr. test (Trials 16–20). These data are shown in Table 4. It may be noted that forgetting tends to increase as degree of learning increases but there is no consistent relationship with simi-

TABLE 3

ANALYSIS OF VARIANCE AMONG EXPERIMENTAL GROUPS ON RECALL OF TASK II AFTER 24 HR.

Source	*df*	Mean Square	*F*
Learning	3	10.58	
Similarity	2	4.62	
L × S	6	28.58	3.73*
Within groups	174	7.66	

* Significant at the 1% level.

TABLE 4

RECALL OF TASK II AFTER 14 MONTHS

Groups	Score		
	Before Rest	After Rest	Difference
Learning			
I	13.09	2.23	10.86
II	13.84	2.98	10.86
III	15.41	2.80	12.61
IV	15.63	2.58	13.05
Similarity			
A	14.95	2.77	12.18
B	14.79	2.52	12.27
C	13.64	2.64	11.00

larity. Analysis of variance on recall scores showed learning significant ($F = 5.50$), similarity not significant ($F = 2.85$), and interaction not significant. Thus forgetting of Task II over 14 months increased as degree of mastery increased. It may be seen from Table 4 that this effect is due to differences among Groups II, III, and IV; there is no difference between Groups I and II.

In view of the finding of significant differences in forgetting over 14 months for the Learning Groups, it is of interest to determine if these groups, as well as the Similarity Groups, differed in performance immediately prior to the 14-month interval. Analysis of variance on the last trial (Trial 20) of the 24-hr. retention test resulted in F's of 6.26 and 4.81 for learning and similarity respectively, both significant at the 1% level. Interaction was not significant. Thus for the four Learning Groups as a whole there were significant differences in performance immediately prior to the long interval, and significant differences in forgetting over the interval, there being a direct relation between performance and forgetting. Since there is no control group, we cannot measure absolute proactive inhibition, but it is clear that there was relative proactive inhibition, resulting from varying amounts of practice on Task I, on retention of Task II after 14 months. We therefore find support for the contention that proactive inhibition, at least relative proactive inhibition, increases as the retention interval increases (6).

Returning to Fig. 1 and 2, we may examine the course of relearning after 14 months. Relearning is relatively rapid, at least more so than original learning of Task I (the control curve for acquisition of Task II). However, the experimental groups begin relearning of Task II at a level of mastery lower than they have ever been on this task, as may be seen by noting the level at which they began *acquisition* of Task II. To state more completely what has occurred: there was relatively large decrement from end of acquisition of Task I to beginning of acquisition of Task II (in the earlier paper it was shown that this decrement increased directly with learning and inversely with similarity). In spite of the decrement, there was positive transfer; all groups began acquisition of Task II at a level above the control. Fourteen months later there was relatively little retention of transfer if we use recall as the measure, i.e., level of mastery on the initial relearning trial. Furthermore, there are obviously no differences among either Learning or Similarity Groups at the beginning of relearning. At this point neither variable was differentially retained. However, if we use relearning as the measure, there is some indication of retention of transfer, as shown by the relatively steep slopes of the curves. And, differences among the relearning curves suggest that there may be differential retention of transfer. This is tested by the following analyses.

The 14-month retention data were first subjected to the usual double classification analysis of variance, in which the score was total correct

responses over all 20 trials. It was found that the interaction mean square closely approached significance, $F = 2.10$, where 2.16 is needed for significance at the 5% level. This may raise some question as to the appropriate error term for testing the main effects. Using within-groups mean square as error, learning was significant at the 5% level ($F = 2.88$), whereas with interaction as error, learning was not significant ($F = 1.37$). Similarity was not significant with either error term.

Even though the above analysis might stand by itself, it was felt that a more sensitive test would be obtained by analyzing for trend. For this, Edwards' (4) analysis of variance for trend was modified to handle double classification.[2] The summary of this analysis is shown in Table 5. The appropriate error term for testing learning, similarity, and their interaction is the mean square between *Ss* in the same group. As may be seen, the *F*'s are the same as those reported above. However, our purpose in performing the trend analysis is to test the interactions of the main variables with trials. For these tests the error term is the mean square of pooled *Ss* × trials. Table 5 shows that interaction of trials and learning is highly significant, but that interaction of trials and similarity, and triple interaction, are not significant. From this analysis we can conclude that degree of learning of Task I is a differentiating variable during relearning of Task II after 14 months, as indicated by the significant *F*'s for learning and for interaction of trials and learning. Similarity, after 14 months, is no longer a significant variable; all degrees of difference between the two tasks are retained and relearned equally well.

[2] Grateful acknowledgment is made to Miss Jean Paulsen, Illinois Neuropsychiatric Institute, Chicago, for developing this method.

TABLE 5

ANALYSIS OF VARIANCE FOR TREND OVER THE 20 RELEARNING TRIALS ON TASK II AFTER 14 MONTHS

Source	*df*	Mean Square	*F*
Learning	3	264.13	2.88*
Similarity	2	33.22	
L × S	6	192.99	2.10
Trials	19	817.71	140.98**
T × L	57	67.06	11.56**
T × S	38	5.29	
T × L × S	114	5.87	1.01
Pooled *Ss* × T	3306	5.80	
Between *Ss* in same group	174	91.73	

* Significant at the 5% level.
** Significant at the 1% level.

Analysis of errors during relearning showed only that Learning Groups exhibiting higher performance in terms of correct responses tended to make *more* errors, whereas for similarity, better-performing groups tended to make fewer errors.

Analysis of time scores gave little reason to think that forgetting of correct responses was merely due to loss of steadiness. Although there was also great forgetting in time scores over 14 months, the correlation between total time and total correct responses over the 20 relearning trials was only .231.

DISCUSSION

Perhaps the most intriguing finding concerns forgetting and relearning over the 14-month period. As indicated by performance on the first trial of relearning after 14 months, there was very great forgetting of Task II, a drop to a level of performance never previously experienced on that task and almost to the level of *Ss* naive to the apparatus. Furthermore, at this recall point, neither the learning nor similarity variables produced differential performance. However, when retention was measured by relearning, 60% of the amount forgotten was regained within 20 trials. The interesting point is that among the Learning Groups latent differences emerged during relearning; originally different levels of mastery

resulted in differential retention. And, it should be added, these levels of mastery resulted, not from varying practice on the transfer task whose retention was being measured, but from different degrees of learning on the training task. Thus we may speak of differential retention of transfer after 14 months without practice.

Although learning as a variable had a strikingly consistent effect throughout all stages of acquisition and retention, differences produced by the kind of formal intertask similarity here employed were not as stable. With successive periods of practice and rest, similarity decreased in effectiveness. At the beginning of practice on the transfer task relatively great differences in performance resulted from varying similarity. However, the differences decreased rapidly so that in the last half of acquisition similarity had a barely significant effect. During relearning after a 24-hr. rest the variable was again effective. After 14 months there was no effect on either recall or relearning of degrees of similarity.

Finally, it may be noted that the results at least partially support the hypothesis that proactive inhibition of retention should increase with length of the retention interval, thereby becoming at least as important a determiner of forgetting as retroactive inhibition (6). Previous studies of proactive inhibition have used retention intervals no longer than a few days. With such intervals it is apparently necessary to use tasks that transfer negatively, either to demonstrate proactive inhibition or that it will become greater than retroactive inhibition (6, 7). The present results show that if the retention interval is long enough, relative proactive inhibition occurs even between tasks that transfer positively.

Summary

With a self-paced discriminative motor task, providing both a set of perceptual-motor paired associates and a steadiness requirement, acquisition and retention of a transfer task were studied as a function of two variables: degree of learning of the training task and similarity between tasks. Since the transfer task was provided by re-pairing the stimuli and responses of the training task, intertask similarity was varied in terms of the number of S–R connections newly paired on the transfer task. After learning both tasks, the transfer task was relearned, once after a 24-hr. interval and again after an average interval of 14 months.

The results were:

1. Transfer was positive and varied significantly both with degree of first-task learning and intertask similarity.
2. After the 24-hr. rest there was some forgetting, measured by recall. There was no evidence for proactive inhibition. In relearning there was proactive facilitation which varied significantly with both main variables.
3. After the 14-month interval forgetting was so great as to result in performance on the recall trial being poorer than that on the first acquisition trial. Performance at recall did not vary either with the degree of learning or similarity of the first task.
4. Relative proactive inhibition occurred over the 14-month interval. The greater the level of mastery immediately prior to the interval, the greater the forgetting. There was no unambiguous relation between forgetting and similarity.
5. Relearning after 14 months was relatively rapid and varied significantly with learning, but not with similarity.

It was pointed out that the results provide some support for the hypothesis that proactive inhibition increases as the retention interval increases.

References

1. Bunch, M. E. The amount of transfer in rational learning as a function of time. *J. comp. Psychol.*, 1936, **22**, 325–337.
2. Bunch, M. E. A comparison of retention and transfer of training from similar material after relatively long intervals of time. *J. comp. Psychol.*, 1941, **32**, 217–231.
3. Duncan, C. P. Transfer in motor learning as a function of degree of first-task learning and inter-task similarity. *J. exp. Psychol.*, 1953, **45**, 1–11.
4. Edwards, A. *Experimental design in psychological research.* New York: Rinehart, 1950.
5. Montgomery, V. E. Transfer in motor learning as a function of distribution of practice. *J. exp. Psychol.*, 1953, **46**, 440–445.
6. Underwood, B. J. Retroactive and proactive inhibition after five and forty-eight hours. *J. exp. Psychol.*, 1948, **38**, 29–38.
7. Underwood, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39**, 24–34.

(Received April 20, 1953)

Journal of Experimental Psychology
Vol. 46, No. 6, 1953

SIMULTANEOUS CONTRAST AS A FUNCTION OF SEPARATION BETWEEN TEST AND INDUCING FIELDS[1]

H. LEIBOWITZ, F. A. MOTE, AND W. R. THURLOW

University of Wisconsin

Variation of stimuli in the visual field while simultaneously investigating the effects of foveal stimulation is a common operation in experimental studies of simultaneous contrast (**6, 9, 10, 22, 24**) and brightness constancy (**7**), and has been shown to be a variable in studies on the absolute threshold (**3, 4, 12**), visual acuity (**8, 18**), flicker (**19**), vernier acuity (**10**), intensity discrimination (**21**), and size constancy (**17**). In order to analyze such interaction in a situation which presents a minimum of parameters, Diamond (**6**) has determined the effect on the binocular visual match between two fields of equal size (designated here as test and reference fields) imaged independently on the two foveas when a third inducing field of the same size is presented adjacent to one of them. He varied the luminance of the inducing and test fields over a wide range. The present study investigates the effect of an additional variable: the separation between inducing and test fields.

Apparatus and Procedure

The basic stimulus pattern presented to the subject is shown in Fig. 1*A*. The fixation point *P* is viewed by both eyes while suitably placed shields restrict vision of the reference field *R* to the left eye and the test and inducing fields, *T* and *I*, to the right eye. The test field is covered by sheet polaroid (J type) so that its luminance can be varied continuously by rotation of an analyzer in front of the right eye. The luminance of all three fields is varied discontinuously by neutral filters (made from photographic film) which are secured to the sheet aluminum from which the stimulus pattern is cut. The light source consists of two 115-v., 150-w. Westinghouse reflector flood lamps connected in parallel and drawing a total of 2.6 amps, DC. The light, after passing through heat-absorbing glass, illuminates evenly a piece of ground-glass located just behind the stimulus pattern. The three fields each subtend a visual angle of 30 min. of arc at a viewing distance of 1.03 m. and are viewed through 2.0-mm. diameter artificial pupils. The fixation point subtends 1.15 min. of arc at the eye and is not varied in luminance. The reference and test fields are both within the limits of the fovea (**20**).

Each experimental session began with ten matches, utilizing the method of adjustment (**14**), between reference and test fields in the absence of the inducing field. The inducing field was then introduced at a value just above threshold and an additional series of ten matches was made between the test and reference fields. This procedure was repeated for successively higher luminance values of the inducing field. An experimental session consisted of matches at ten luminance values of the inducing field from 0 to 400 millilamberts in steps of approximately .3 to .5 log units at a fixed separation between inducing and test fields. Stimulus patterns with angular separations between the closest edges of test and inducing fields of 0′, 10′, 30′, 60′, 180′, and 540′ were employed. The luminance of the

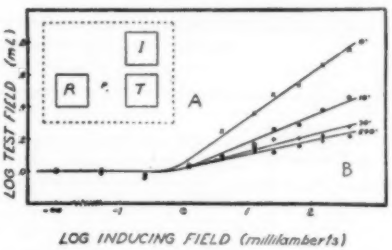

FIG. 1. (*A*) Diagram of the stimulus pattern seen by *S* for 10′ separation. (*B*) The luminance of the test field required for a constant visual effect for various luminance values of the inducing field and for the indicated separations between test and inducing fields.

[1] This research was supported by funds provided by the State Legislature.

reference field was maintained at 1.0 millilambert throughout the experiment.

Four *Ss* were used, three of whom (T, M, and C) had no previous experience in visual experimentation. Subject L had previously served in the Diamond study (6).

Results

Table 1 gives the average luminance of the test field required to match the reference field viewed by the other eye for ten luminance values of the inducing field and for six separations between test and inducing fields. Each datum is the average of three sessions for Subject T and two sessions for the other *Ss*. The means for all *Ss* at separations of 0′, 10′, 30′, and 540′ are plotted in Fig. 1*B*. Because of overlap, the curves for the 60′ and 180′ separations have not been plotted.

When the luminance of the inducing field is less than the test field, there is little or no effect upon the luminance of the test field required to produce a constant visual effect, i.e., match the fixed reference field. This is true for all degrees of separation between test and inducing fields. When the inducing field luminance is greater than that of the test field, additional test field luminance is required for the constant effect. The additional luminance required is proportional, on a double log plot, to the luminance of the inducing field and decreases with increasing separation between test and inducing fields.

When test and inducing fields are contiguous (zero separation), the data are comparable to those obtained by Diamond, although his experimental procedure differed from that reported here. He varied the luminance of the field presented alone to the left eye to match the test field as seen by the right eye which was viewed simultaneously with the inducing field. The data indicate no essential difference as a result of the different procedures employed by Diamond and in the present study. In both experiments, the effectiveness of foveal stimulation is decreased when an adjacent inducing field of the same size and greater luminance is present, the decrease in effectiveness being proportional (on a double log grid) to the inducing field luminance.

The effect of separating the inducing and test fields can be determined by comparison of the rising portions of the curves in Fig. 1*B*. With increased

TABLE 1

Log Test Field Luminance Required to Match a 1.0 Millilambert Reference Field for Various Log Luminance Values and Separations from an Equal-Sized Inducing Field

Log Inducing Field Luminance	Separation					
	0′	10′	30′	60′	180′	540′
$-\infty$	.00	.00	.00	.00	.00	.00
−1.30	.01	−.01	−.01	.01	.04	.03
−0.61	−.03	−.04	.00	.04	.08	−.02
0.08	.03	.04	.04	.07	.08	.04
0.61	.25	.10	.08	.09	.10	.09
1.12	.36	.17	.13	.12	.17	.15
1.42	.48	.26	.20	.08	.19	.12
1.82	.54	.29	.21	.12	.20	.16
2.18	.66	.38	.19	.12	.20	.22
2.60	.75	.46	.28	.20	.24	.22

separation, there is a decrease in slope reflecting a diminution in the inhibitory effect produced by a given inducing field luminance. This diminution is greatest, for a given amount of separation, the nearer the inducing and test fields are to each other. For separations greater than 30′, i.e., when the inducing field is entirely outside the fovea, no consistent trends are observable in either the individual or group data. There is a residual effect for separations greater than 30′ but no systematic variation in its magnitude as a function of separation was observed in the present study.

Discussion

The results of the present study extend Diamond's conclusions, obtained with adjacent test and inducing fields, to include conditions where the equal-sized fields are separated. These conclusions are: (*a*) for inducing field luminances less than the test field, little or no effect upon the test field is produced, and (*b*) for inducing field luminance values greater than the test field, the latter is inhibited in direct proportion (on a double log plot) to the luminance of the inducing field. The effect of separation between test and inducing fields is to decrease the inhibitory effect of the inducing field.

The present data and those of Diamond are consistent with the conclusion of Graham and Granit (15) and Graham (13) that an inhibitory mechanism exaggerates differences in the luminances of objects imaged on the fovea. Graham (13) has presented evidence suggestive of a neural basis for such interaction; Shouten and Ornstein (22) have hypothesized that electric potentials generated by the inducing field diminish the sensitivity of the entire retina; and Fry (9) has reviewed these and other mechanisms which might account for the phenomenon under discussion. The present data do not provide a conclusive basis for choosing among the various inhibitory mechanisms which could be effective in the fovea. With respect to peripheral inducing fields, however, it may be significant that most of the observed inhibitory effect occurs when the inducing field is wholly or partially within the fovea, and that for inducing fields beyond the fovea the residual effects are variable from *S* to *S* and are not affected systematically by the position of the inducing field. Fry and Alpern (11) have demonstrated that the "veiling glare" produced by a large (9.5° in diameter) field of low luminance superimposed on the test field can produce an inhibitory effect in the fovea similar to that from a small (2.5° by .5°) inducing field of high luminance imaged in the peripheral retina. Thus, the inhibitory effect of peripheral inducing fields is shown to be similar to that produced by a "veil" of light such as would be expected to result from stray light falling on the retina. On this basis, Fry (9) suggests that stray light acting at the borders of the test field is the most probable mechanism by which peripheral inducing fields can modify the effects of foveal stimulation. The prevalence of stray light in the eye (2, 5), evidence that peripheral retinal interaction in general produces summation rather than inhibition (16), and the observation that only residual effects were obtained from peripheral inducing fields in the present study are consistent with this conclusion.

These data represent quantitatively the phenomenon of simultaneous brightness contrast. Since these functions were obtained by systematic variation of variables over a wide range, they should be useful as a base for quantitative interpretation of visual data, such as constancy, obtained under conditions where visual stimuli, in addition to the foveal test stimulus, coexist in the visual field.[3] The potential value of such interpretation will be enhanced by extending the present type of analysis to include additional parameters such as area, the extent to which the inducing field surrounds the test field, temporal factors (metacontrast: 1), and wave-length distribution (chromatic contrast: 13, 23).

[3] The essence of this analysis was suggested by Prof. C. H. Graham, personal communication, 1952.

Summary

By means of a binocular matching technique the interaction between a foveal test field and an equal-sized inducing field was investigated for various separations between the two fields.

For all separations, little or no effect is produced in the test field when the luminance of the inducing field is less than that of the test field. For inducing field luminances greater than the test field, the test field is inhibited in direct proportion, on a double log plot, to the luminance of the inducing field.

The present data are consistent with the assumption that luminance differences in the fovea are exaggerated by an inhibitory mechanism. The ability of inducing fields imaged in the peripheral retina to inhibit activity in the fovea is apparently due to stray light.

The relevance of the present data to constancy and other forms of retinal interaction, as well as contrast, is suggested.

References

1. Alpern, M. Metacontrast. *Amer. J. Optom. and Arch. Amer. Acad. Optom.*, 1952, **29**, 631–646.
2. Bartley, S. H. *Vision: a study of its basis.* New York: Van Nostrand, 1941.
3. Beitel, R. J. Spatial summation of subliminal stimuli in the human retina. *J. gen. Psychol.*, 1934, **10**, 311–327.
4. Beitel, R. J. Inhibition of threshold excitation in the human eye. *J. gen. Psychol.*, 1936, **14**, 31–61.
5. Boynton, R. M., & Riggs, L. A. The effect of stimulus area and intensity upon the human retinal response. *J. exp. Psychol.*, 1951, **42**, 217–226.
6. Diamond, A. L. Foveal simultaneous brightness contrast as a function of inducing- and test-field luminances. *J. exp. Psychol.*, 1953, **45**, 304–314.
7. Feldman, S., & Weld, H. P. Perceiving. In E. G. Boring, H. S. Langfeld, and H. P. Weld (Eds.), *Psychology.* New York: Wiley, 1935.
8. Fisher, B. M. The relationship of the size of the surrounding fields to visual acuity in the fovea. *J. exp. Psychol.*, 1938, **23**, 215–238.
9. Fry, G. A. Mechanisms subserving simultaneous brightness contrast. *Amer. J. Optom. and Arch. Amer. Acad. Optom.*, 1948, **25**, 162–178.
10. Fry, G. A. Effects of uniform and non-uniform surrounds on foveal vision. *Amer. J. Optom. and Arch. Amer. Acad. Optom.*, 1950, **27**, 423–436.
11. Fry, G. A., & Alpern, M. The effect of a peripheral glare source upon the apparent brightness of an object. *J. opt. Soc. Amer.*, 1953, **43**, 189–195.
12. Geldard, F. A. Brightness con ast and Heyman's law. *J. gen. Psychol.*, 1931, **5**, 191–206.
13. Graham, C. H. Vision: III. Some neural correlations. In C. Murchison (Ed.), *Handbook of general experimental psychology.* Worcester, Mass.: Clark Univer. Press, 1934.
14. Graham, C. H. Behavior, perception, and the psychophysical methods. *Psychol. Rev.*, 1950, **57**, 108–120.
15. Graham, C. H., & Granit, R. Comparative studies on the peripheral and central retina: VI. Inhibition, summation, and synchronization of impulses in the retina. *Amer. J. Physiol.*, 1931, **98**, 664–673.
16. Granit, R. *Sensory mechanisms of the retina.* London: Oxford Univer. Press, 1947.
17. Holway, A. H., & Boring, E. G. Determinants of apparent size with distance variant. *Amer. J. Psychol.*, 1941, **64**, 21–37.
18. Lythgoe, R. J. *The measurement of visual acuity.* London: His Majesty's Stationery Office (Spec. Rep. Ser., No. 173), 1932.
19. Lythgoe, R. J., & Tansely, K. *The adaptation of the eye: its relation to the critical frequency of flicker.* London: His Majesty's Stationery Office (Spec. Rep. Ser., No. 134), 1929.
20. Polyak, S. L. *The retina.* Chicago: Univer. of Chicago Press, 1941.
21. Ratoosh, P., & Graham, C. H. Areal effects in foveal brightness discrimination. *J. exp. Psychol.*, 1951, **42**, 367–375.
22. Shouten, J. F., & Ornstein, L. S. Measurements on direct and indirect adaptation by means of a binocular method. *J. opt. Soc. Amer.*, 1939, **29**, 168–182.
23. Troland, L. T. Vision: I. Visual phenomena and their stimulus correlations. In C. Murchison (Ed.), *Handbook of general experimental psychology.* Worcester, Mass.: Clark Univer. Press, 1934.
24. Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received March 23, 1953)

Journal of Experimental Psychology
Vol. 46, No. 6, 1953

THE LOCUS OF SHORT DURATION AUDITORY FATIGUE OR "ADAPTATION"

J. DONALD HARRIS AND ANITA I. RAWNSLEY

U. S. Naval Medical Research Laboratory

Auditory fatigue has roughly been defined as a reversible shift in threshold after acoustic stimulation. However, when one examines the facts closely it appears that at least two types of reversible threshold shift may be found which can best be described as auditory fatigue and auditory adaptation. The term "adaptation" was proposed simultaneously by Caussé and Chavasse (2) and Lüscher and Zwislocki (18). In general, it is thought of as arising from stimulus intensities and durations too weak and brief to produce auditory fatigue; at least, the resulting threshold shift is too transitory to measure with the usual audiometric techniques.

It is possible that the two phenomena may be different manifestations of the same reversible process, though this seems unlikely in view of the fact that adaptation can be distinguished from fatigue by any of four criteria (8): (*a*) In fatigue, the duration of the stimulating tone has a cumulative effect from 30 sec. up to at least 10 min. (12, 14), while in adaptation the effect of duration of stimulation is not cumulative within the rather wide limits of .05 sec. up to at least 10 sec. (9, 16, 20). (*b*) In adaptation, a plot of the auditory threshold as it returns to normal after termination of the adapting stimulus is a straight line (18, 19, 20), while in even very mild fatigue the temporal course of recovery of threshold shows a much more pronounced negative acceleration (12, 13). (*c*) In adaptation, the frequency region of maximum threshold shift is at the stimulus frequency (2, 4, 14, 18, 21), whereas in fatigue the maximum effect may lie a half-octave higher (3). (*d*) In fatigue, the shape of the recruitment curve is a straight line (11), while, as we shall see in this paper, in adaptation the recruitment data form a curve of quite definite characteristics.

For these reasons adaptation must be considered as at least a special case of fatigue, if it is not indeed a separate phenomenon altogether. Because the threshold shift is constant through stimulus durations from .05 to at least 10 sec., adaptation must be thought of as a changed dynamic equilibrium set up somewhere within the auditory system.

If, as we now believe, it is necessary to make a preliminary differentiation between fatigue and the more recently discovered adaptation, we may now inquire where in the auditory system is the locus of adaptation. A search for this locus is the aim of the present paper.

The site of adaptation is not the middle ear, since we have found (20) that a patient with a radical ossilectomy yielded normal curves of recovery from adaptation. Lüscher and Zwislocki (19) by experiments on cross-conduction have ruled out the central nervous system. By elimination, then, the locus is evidently placed within the cochlea. It is not clear, however, whether neural or nonneural structures within the cochlea are responsible. Rosenblith (22) and Rosenzweig and Rosenblith (23) held that neural mechanisms were responsible for a diminished neural response to the second of two clicks. But their argu-

ment is not quite conclusive. An acoustic stimulus was used which set into play the whole auditory system, and consequently a diminished neural response might be recorded as a result of an adaptation process at the organ of Corti, or the transmission mechanism between the organ of Corti and the first-order neuron.

An additional line of attack has recently become possible with the demonstration (**5, 6, 24**) that the phenomenon of loudness recruitment is distinctly nonneural, and is indeed probably confined to the hair cells of the organ of Corti (for summary see **10**). It should be possible, now, to examine whether an ear in a state of adaptation exhibits recruitment. If it does, we have inferential evidence that the state of adaptation is intimately connected with the organ of Corti.

The present study was designed to seek for recruitment in an ear already thrown into a state of adaptation by a preceding tonal stimulus. The loudness of a brief probe tone in the adapted ear was adjusted to match the loudness of an identical simultaneous tone in the nonadapted ear. From these binaural loudness balances at a variety of loudness levels it was possible to explore for the presence of loudness recruitment in the ear exhibiting adaptation.

Apparatus and Procedure

The output of a General Radio Oscillator at 1000 cps was split and led through the two channels of an amplifier which shaped the envelopes to produce rise-fall times of 25 msec. each. An electronic timer governed the temporal pattern of stimuli. Figure 1 illustrates the instrumentation and channels. One channel, Tone A, 450 msec. in duration, was led to one of a pair of earphones, through the appropriate filtering, attenuation, and impedance matching. The other channel was filtered and split into two channels, Tones B and C, both 50 msec. in duration, one going to each ear. By this means *E* was enabled to throw one ear into a state of adaptation with Tone A, and subsequently to ask for binaural loudness matches between the simultaneous Tones B and C. The actual steps used were as follows:

1. Threshold was found for Tone B alone.
2. Tone B was raised 10, 20, or 30 db over threshold.
3. With Tone B at one of these new levels, Tone A was introduced and raised until the resulting adaptation rendered Tone B just audible. These new thresholds for Tone B indicated that the ear was at that moment at a level of adaptation amounting to 10, 20, or 30 db.
4. Tone C was set at either 16, 100, 360, 1000, or 2200 millisones.
5. The full three-tone sequence was presented as in Fig. 1. Tone B was adjusted from its new threshold by *S* to produce equal loudness with Tone C at whichever of the five reference loudnesses Tone C was used to produce.

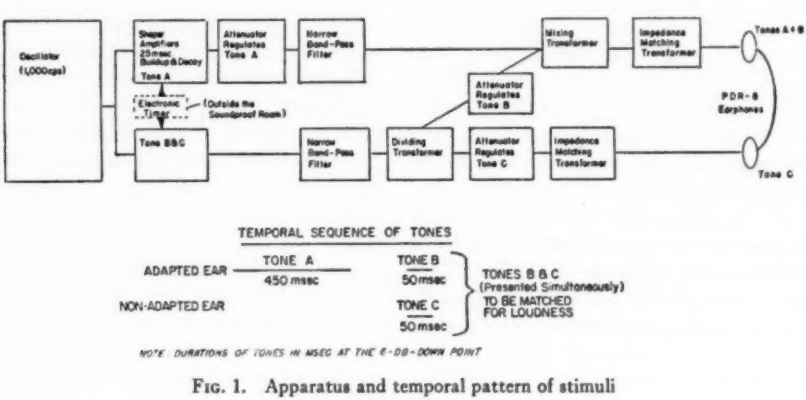

Fig. 1. Apparatus and temporal pattern of stimuli

Each of eight normal-hearing *Ss* made five or more binaural loudness matches at each experimental condition. Since five reference loudnesses were used at each of the three levels of adaptation, there were 15 experimental conditions.

Results

Each point on the curves of Fig. 2 represents the mean loudness match for one combination of level adaptation and one reference loudness. On the ordinate are the reference loudnesses for Tone C in the normal, or nonadapted ear. On the abscissa are the intensity levels of Tone B which furnished the loudness matches for Tone C. These levels of Tone B are expressed in terms of decibels above threshold for Tone B. Note that the three curves intersect the abscissa at exactly 10, 20, or 30 db; this is because (Step 3 above) Tone A was adjusted until threshold for Tone B was shifted by one of these values.

The curves may be read as follows: In Curve 2, the threshold for Tone B when presented alone was shifted 20 db when preceded by Tone A; when Tone C was set at 100 millisones (20 db above threshold) Tone B had to be raised only about 10 db above its shifted threshold to achieve equal loudness. In other words, raising the tone 10 db in the adapted ear was equivalent to raising the same tone 20 db in the nonadapted ear. This faster growth of loudness in the ear listening to Tone B indicated the presence of recruitment in that ear.

The more closely the curves in Fig. 2 approach the 45° diagonal, the more nearly complete is recruitment said to become. In Curve 1, for example, a reference loudness of 2200 millisones (50 db over threshold for Tone C) is equalled by raising Tone B 50 db over threshold also. Thus, the 10-db threshold loss of Tone B is overcome and recruitment is complete.

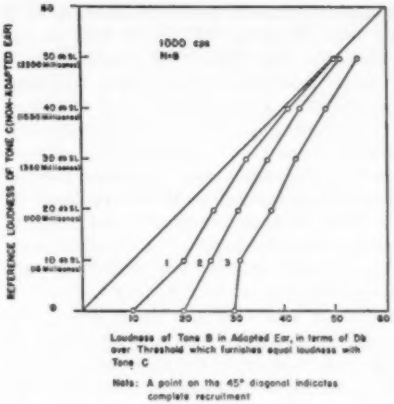
REFERENCE LOUDNESS OF TONE C (NON-ADAPTED EAR)
1000 cps
N=8
1
2
3
Loudness of Tone B in Adapted Ear, in terms of Db over Threshold which furnishes equal loudness with Tone C
Note: A point on the 45° diagonal indicates complete recruitment

Fig. 2. Shows recruitment in auditory adaptation

In Fig. 2, all three curves reach or tend to reach the 45° diagonal, and the interpretation is that, at all levels of adaptation used here, pronounced recruitment occurs.

Discussion

The locus of adaptation.—From the presence of recruitment in the adapted ear, both being simultaneous and transient phenomena, one can infer that recruitment and adaptation are intimately related, and are most probably expressions in one and the same tissues of reversible aftereffects of stimulation. Hood (14) discussed an inference that "adaptation" and recruitment are closely related, when he found that an ear suffering from Ménière's disease (and therefore most probably of a recruiting type) exhibits one feature (named "relapse" by Hood) similar to that found in an ear undergoing "adaptation." Hood, however, uses the term "adaptation" to apply to his "per-stimulatory fatigue," a slow cumulation of loss of loudness extending over several minutes of continuous stimulation. It is not at all clear that per-stimulatory fatigue, and what we refer to as adaptation in this paper, are one and the same thing. Most probably they are not, since what we call adaptation is a

process complete within a fraction of a second. Huizing (15) found that an ear exhibiting recruitment by reason of perceptive deafness suffers more auditory fatigue than a normal ear, and of course recruitment in auditory fatigue has often been demonstrated (1, 3, 11, 17). These workers did not indeed separate fatigue from adaptation, but there are hints in de Maré (4) that a recruiting ear exhibits more adaptation than a normal ear. When these studies are added to our demonstration of recruitment and adaptation at one and the same time, the weight of evidence points to the locus of adaptation as being the same as the locus of recruitment, namely, in the end-organ itself, not the nervous system.

The shape of the recruitment curve in an adapted ear.—A sharp distinction is found in the growth of loudness in a fatigued and in an adapted ear. It is seen in Fig. 2 that the curve of recruitment is one of negative acceleration (the so-called "asymptotic" type, 10). In the case of a fatigued ear, on the other hand, the curve of recruitment is usually one of positive acceleration or of the straight-line type (3, 11). We have here further evidence that adaptation and fatigue must be distinguished. It is more than likely that future clinical work, looking toward discovering the types of deafness which produce different shapes of recruitment curve, may use this curve to assist in determining more precisely the tissues involved in fatigue and in adaptation.

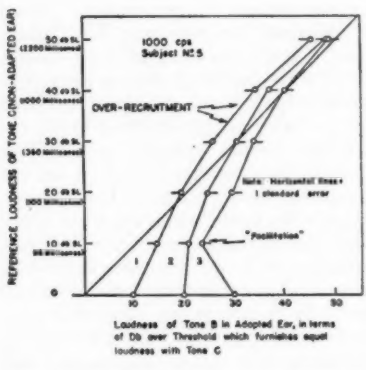

FIG. 3. Shows recruitment in auditory adaptation

Overrecruitment and "facilitation."—Two of our *S*s showed overrecruitment, a phenomenon sometimes encountered in unilateral perceptive deafness, in which with equal intensity input to the two ears, the loudness in the defective ear is even *greater* than in the normal ear. This condition is shown for one *S* in Fig. 3 (those points plotted above the 45° diagonal). The only explanation given in the literature for overrecruitment is in terms of a diminution of central inhibition consequent on a reduction in number of available nerve fibers or end-organ units (7, p. 586). This explanation could serve in the present case if one assumes a reduction in the number of *functionally active* end-organ units.

Four of our eight *S*s showed a phenomenon contrary to expectation, in which Tone B, originally at threshold, leaped into audibility when Tone C was added simultaneously and in phase in the other ear. This enhancement of loudness we call here "facilitation." The most extreme case of facilitation is seen in Curve 3 of Fig. 3. Three other *S*s showed enhancements of loudness amounting to 2–6 db. Undoubtedly the phenomenon is related in some way to observations of super-normal periods of nerve response following acoustic stimulation, and to observations of changes in sensitivity as a function of phase relations.

SUMMARY

Auditory adaptation, defined as a reversible shift in threshold following very mild acoustic stimulation, was distinguished from true auditory fatigue by four criteria, one of which was the shape of the curve of recruitment.

This experiment sought to throw light upon where, in the auditory system, adaptation occurs. By the technique of simultaneous binaural loudness balancing between one normal ear and another ear undergoing a threshold shift as a result of a preceding adapting stimulus, it was possible to demonstrate an accelerated growth of loudness (recruitment) in the adapted ear. The intimate temporal relation between adaptation and recruitment, the latter distinctly allied to the organ of Corti, was thought to point to the

peripheral organ, not the nervous system, as the site of the recently discovered phenomenon of auditory adaptation.

References

1. Békésy, G. v. Zur Theorie des Hörens. *Phys. Z.*, 1929, **30**, 721–745.
2. Caussé, R., & Chavasse, P. Études sur la fatigue auditive. *Ann. Psychol.*, 1947, **43–44**, 265–298.
3. Davis, H., Morgan, C. T., Hawkins, J. E., Jr., Galambos, R., & Smith, F. W. Temporary deafness following exposure to loud tones and noise. Report No. PBM 57250, Comm. Med. Sci., O.S.R.D., 30 Sept. 1943. Available from Office of Technical Services, U. S. Dept. of Commerce, Washington, D. C. Sharply abridged in: *Laryngoscope, St. Louis*, 1946, **56**, 19–21, and in *Acta Oto-laryngol., Stockh.*, 1950, Suppl. 88, p. 57.
4. De Maré, G. Audiometrische Untersuchungen. *Acta Oto-laryngol., Stockh.*, 1939, **31**, 1–173.
5. Dix, M. R., Hallpike, C. S., & Hood, J. D. Observations upon the loudness recruitment phenomenon, with especial reference to the differential diagnosis of disorders of the internal ear and VIII nerve. *Proc. roy. Soc. Med.*, 1948, **41**, 516–526.
6. Eby, L. G., & Williams, H. L. Recruitment of loudness in the differential diagnosis of end-organ and nerve fiber deafness. *Laryngoscope, St. Louis*, 1951, **61**, 400–414.
7. Fowler, E. P. The use of threshold and louder sounds in clinical diagnosis. *Laryngoscope, St. Louis*, 1938, **48**, 572–588.
8. Harris, J. D. The several short term effects following "mild" acoustic stimulation. *J. acoust. Soc. Amer.*, 1953, **25**, 187. (Abstract)
9. Harris, J. D. Studies in short-duration auditory fatigue: II. The effect of the duration of the stimulating tone. III. The effect of the interval between stimuli. U. S. Navy Med. Res. Lab. Rep. No. 168, 8 Dec. 1950.
10. Harris, J. D. A brief critical review of loudness recruitment. *Psychol. Bull.*, 1953, **50**, 204–220.
11. Harris, J. D., Haines, H. L., & Myers, C. K. Loudness perception for pure tones and for speech. *Arch. Oto-laryngol., Chicago*, 1952, **55**, 107–133.
12. Harris, J. D. Recovery curves and equinoxious exposures in reversible auditory fatigue following stimulation up to 140 db. *Laryngoscope*, in press.
13. Hirsh, I. J., & Ward, W. D. Recovery of the auditory threshold after strong acoustic stimulation. *J. acoust. Soc. Amer.*, 1952, **24**, 131–141.
14. Hood, J. D. Studies in auditory fatigue and adaptation. *Acta Oto-laryngol., Stockh.*, 1950, Suppl. 92. Summarized in Hallpike, C. S., & Hood, J. D. Some recent work on auditory adaptation and its relation to the loudness recruitment phenomenon. *J. acoust. Soc. Amer.*, 1951, **23**, 270–274.
15. Huizing, H. C. The relation between auditory fatigue and recruitment. *Acta Oto-laryngol., Stockh.*, 1948, Suppl. 78, 169–172.
16. Kelsey, P. A., & Rawnsley, A. I. Adaptation of the ear to sound stimuli: the intensity-time relationship. U. S. Navy Med. Res. Lab. Rep. No. 218, 26 Oct. 1952.
17. Larsen, B. Undersogelser over Lorelsens traethed. *Nordisk Medicin*, 1942, **14**, 1714–1718.
18. Lüscher, E., & Zwislocki, J. The decay of sensation and the remainder of adaptation after short pure-tone impulses on the ear. *Acta Oto-laryngol., Stockh.*, 1947, **35**, 428–445.
19. Lüscher, E., & Zwislocki, J. Adaptation of the ear to sound stimuli. *J. acoust. Soc. Amer.*, 1949, **21**, 135–139.
20. Rawnsley, A. I., & Harris, J. D. Studies in short-duration auditory fatigue: II. Recovery time. *J. exp. Psychol.*, 1952, **43**, 138–142.
21. Rawnsley, A. I., & Harris, J. D. Patterns of cochlear adaptation at three frequency regions. *J. acoust. Soc. Amer.*, in press.
22. Rosenblith, W. A. Auditory masking and fatigue. *J. acoust. Soc. Amer.*, 1950, **22**, 792–800.
23. Rosenzweig, M. R., & Rosenblith, W. A. Some electro-physiological correlates of the perception of successive clicks. *J. acoust. Soc. Amer.*, 1950, **22**, 878–880.
24. Saltzman, M. Recruitment phenomenon in craniocerebral trauma. *Arch. Oto-laryngol., Chicago*, 1950, **52**, 782–785.

(Received March 16, 1953)

ERRATUM

JANET A. TAYLOR AND KENNETH W. SPENCE
Northwestern University *State University of Iowa*

In reporting the procedure employed in a serial, verbal maze learning experiment published in this Journal, 1952, **44**, 61–64, an unfortunate error was made in describing the exposure interval. The method used was identical with that employed in the standard rote serial learning situation, each unit of the series being exposed for 2 sec. with the anticipation of the succeeding unit during the 2-sec. presentation of the just preceding unit. There was not a 2-sec. interval between exposures of each unit as was stated in the article. A 4-sec. interval would have allowed too much time for such uncontrollable factors as rehearsal or thinking ahead in such a simple learning situation.

(Received October 16, 1953)

A clear and comprehensive exposition of the revolutionary "Q-Technique" and its methodology.

The Study of Behavior

By WILLIAM STEPHENSON

In this long-awaited work, Professor Stephenson redefines the scope of behaviorism from the standpoint of modern logical analysis to include both internal and external frames of reference. He applies his principles to many branches of psychology, showing the ultimate extensions of the theory into the social sciences and into many aspects of the humanities. *$7.50*

At your bookstore, or from

THE UNIVERSITY OF CHICAGO PRESS

5750 Ellis Ave., Chicago 37, Ill.

PHILOSOPHICAL LIBRARY PUBLICATIONS

SCIENCE and MAN'S BEHAVIOR

The Contribution of Phylobiology

by TRIGANT BURROW, M.D., Ph.D.

A VITAL EXCHANGE OF VIEWS ON THE PROBLEM OF HUMAN CONFLICT

Dr. Burrow clarifies fundamental aspects of his altered orientation developed in a lifetime of research. The findings of his pioneer studies of group behavior are discussed with:

P. W. BRIDGMAN
A. J. CARLSON
JOHN DEWEY
KURT GOLDSTEIN
C. JUDSON HERRICK
LELAND E. HINSIE
CLYDE KLUCKHOHN
GARDNER MURPHY

and other men of science.

Includes the full text of *The Neurosis of Man* on which the discussion is based. Edited by William E. Galt, Ph.D., *Research Associate, The Lifwynn Foundation.*

$6.00

A HANDBOOK of MUSIC THERAPY

Edited by
EDWARD PODOLSKY, M.D.

The latest facts of music therapy are presented in a series of articles by the leading psychiatrists, music therapists and clinical psychologists—based on clinical experience with music in a variety of ailments. $6.00

EXISTENTIAL PSYCHOANALYSIS

by JEAN-PAUL SARTRE

Sartre here blends philosophy and psychology in presenting a new psychoanalysis based on the principles of existentialism. He criticizes psychoanalytical schools and attempts to justify a method which interprets the human personality in terms of a fundamental choice of being. $4.75

01849

PHILOSOPHICAL LIBRARY, PUBLISHERS

15 East 40th Street Desk 188 New York 16, N.Y.

Expedite shipment by prepayment.